Causes and Consequences of Recombination Rate Variation

by

Caitlin Suzanne Smukowski Heil

Department of Biology
Duke University

Date:_____

Approved:

_____
Mohamed Noor, Supervisor

_____
John Willis

_____
Storrs Mitchell-Olds

_____
Thomas Petes

_____
Corbin Jones

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2014

ABSTRACT

Causes and Consequences of Recombination Rate Variation

by

Caitlin Suzanne Smukowski Heil

Department of Biology
Duke University

Date:_____

Approved:

_____
Mohamed Noor, Supervisor

_____
John Willis

_____
Storrs Mitchell-Olds

_____
Thomas Petes

_____
Corbin Jones

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology in the Graduate School of Duke University

2014

Abstract

Recombination is the process in which genetic material is exchanged between one's homologous chromosome pairs during egg or sperm development (meiosis). Recombination is necessary for proper segregation of chromosomes during meiosis, and also plays a role in purging deleterious mutations, accelerating adaptation, and influencing the distribution of genomic features over evolutionary time. While recombination is clearly an important process, recombination rate is known to vary within and between individuals, populations, and species. Furthermore, what causes this variation remains relatively unknown. Using empirical and sequenced based estimates of recombination rate for the closely related species *Drosophila pseudoobscura* and *Drosophila miranda*, I seek to understand where recombination happens across the genome, to what extent recombination changes between species, and what genomic features are responsible for these changes. These data will deepen our understanding of mechanisms determining the recombination landscape, and shed light on generalized patterns and exceptions of recombination rate variation across the tree of life.

Contents

Abstract iv

List of Tables viii

List of Figures ix

Acknowledgements x

1. Recombination rate variation in closely related species 1

Introduction 1

Determinants and Correlates 4

Molecular Evolutionary Consequences of Recombination Rate Variation 8

Conservation and Divergence of Recombination 9

How does the methodology by which recombination is measured affect estimates of recombination rate? 10

Why would we see conservation of recombination rates between some species and not others? 13

Why would we see recombination rate conservation at some scales and not others? 15

Should we expect to see conservation between closely related species? 18

Conclusions 20

2. Recombination modulates how selection affects linked sites in *Drosophila* 33

Smukowski Heil Contribution 33

Introduction 33

Materials and Methods 40

Results 49

General Summary of Recombination Data: Fine-Scale Maps 49

General Summary of Recombination Data: Ultrafine-Scale Maps 53

Recombination Rate Comparison between Maps 54

Diversity, Divergence, and Recombination 59

Recombination and Selection 63

Discussion 69

3. Zinc finger binding motifs do not explain recombination variation within or between species of Drosophila 78

Introduction 78

Materials and Methods 83

Results 89

Prdm9 homology 89

Candidate protein analysis 91

A comprehensive search for a zinc finger binding sequence motif 96

Differences in motif occurrences do not account for changes in recombination landscapes between closely related species 97

Sequence motifs 99

A validation of our approach using human recombination data 100

Discussion 100

4. No detectable effect of DNA methyltransferase DNMT2 on meiotic recombination . 106

Introduction 106

Materials & Methods 110

Results 115

Recombination across a euchromatic region 115
Discussion 116
5. The fine scale recombination landscape in two species of Drosophila 120
Introduction 120
Materials and Methods 124
Results and Discussion 128
LD-based recombination maps are good estimates of empirical recombination maps 128
Between species recombination rate comparison 130
Distribution of recombination rate across the genome 133
Conclusions 137
Works Cited 138
Biography 175

List of Tables

Table 1: Relationships between recombination rate and measures of diversity and divergence 23

Table 2: Relationships between recombination rate and measures of diversity and divergence 27

Table 3: A comparison of recombination rates between species pairs 31

Table 4: Comparison of intervals condensed within and between recombination maps. 57

Table 5: Factors affecting diversity within species at four-fold degenerate sites for unpreferred codons using intervals with conserved recombination rate 61

Table 6: Factors affecting divergence between species at four-fold degenerate sites for unpreferred codons using intervals with conserved recombination rate 61

Table 7: Test for relationship between recombination rate and number of nonsynonymous substitutions 65

Table 8: PRDM9 candidate proteins 90

Table 9: Proteins with motifs significantly associated with recombination 94

Table 10: Markers used to assay recombination 114

List of Figures

Figure 1: Expectation of the relationship between divergence and recombination rate if the recombination–diversity positive correlation is the result of recombination being mostly mutagenic or the result of recombination's effect on selection at linked sites. 37

Figure 2: Relationships of study species 39

Figure 3: Fine-scale recombination rates on chromosome 2 52

Figure 4: Ultrafine recombination rates 54

Figure 5: Fine-scale recombination rates for condensed intervals without and with global modifier correction 55

Figure 6: Relationship of recombination rate to diversity (filled circles, solid line) and divergence (open circles, dotted line) for fine-scale regions with conserved recombination between D. pseudoobscura–D. miranda. 63

Figure 7: Footprints of diversity around substitutions 68

Figure 8: Footprints in diversity around substitutions 69

Figure 9: A model of Cys2His2 zinc finger binding 80

Figure 10: Predictions to test if changes in DNA binding motifs between species account for variation in recombination rate between species 98

Figure 11: DNA methylation crossing scheme 111

Figure 12: Comparison of *D. pseudoobscura* empirical and LDhelmet recombination rates on chromosome 2 129

Figure 13: Fine scale recombination rates (10kb) between species 133

Figure 14: Recombination rate around transcription start sites 134

Figure 15: Recombination rate within genes 135

Figure 16: Proportion recombination occurring in proportion of sequence 136

Acknowledgements

I would like to acknowledge my advisor, Mohamed Noor, for his unparalleled guidance, assistance, positive attitude, and friendship throughout my graduate career. I would also like to thank members of the Noor lab past and present, in particular Brenda Winkler, Laurie Stevison, and Suzanne McGaugh. Undergraduate students Bret Lesavoy and Leigh Raskauskas helped me with genotyping for Chapter 4, for which I'm grateful. I would also like to acknowledge my committee: John Willis, Tom Mitchell-Olds, Tom Petes, and Corbin Jones for their comments, advice, and support. Duke University Biology department is a great place to work because of the Biology department support staff, in particular Randy Smith, Jo Bernhardt, Anne Lacey, Jim Tunney, and Ryan Hardy. This work would not be possible without funding provided by Duke Biology, James B. Duke fellowship, Sigma Xi, Society for the Study of Evolution, National Institutes of Health, and the National Science Foundation. I am especially indebted to my fellow Biology graduate students for making my time at Duke collaborative, productive, thought-provoking, and most of all, happy. Finally, I'd like to thank my parents Patty and David, my sister Anna, and my husband Lou for their love and encouragement.

1. Recombination rate variation in closely related species

Caiti Smukowski and Mohamed Noor

Citation: Smukowski, C. S. and M. A. F. Noor (2011). "Recombination rate variation in closely related species." Heredity **107**(6): 496-508.

Introduction

One of the main goals of evolutionary genetics is to understand how variation is created and maintained within and between species. Homologous meiotic recombination affects variation in the genome: differences in the rate of recombination determine the amount of genetic variation within populations and the rate at which new combinations of alleles are introduced into populations (Brooks et al. 1986; Coop et al. 2007). Indeed, in flies and humans, variability in recombination rate explains more than 50% of variation in nucleotide heterozygosity across the genome (Nachman 2002), and recombination may shape features of the genomic landscape such as codon bias, base composition, and the distribution of repetitive elements and polymorphisms ((Charlesworth et al. 1994; Comeron et al. 1999; Duret et al. 2008), see also section "Molecular Evolutionary Consequences of Recombination Rate Variation"). Recombination rates are also expected to mediate the effectiveness of natural and sexual selection on genome evolution because the breakdown of linkage between nucleotide sites allows sites to behave independently, permitting selection to act efficiently,

eliminate mutations, reduce genetic hitchhiking, and facilitate adaptive evolution by reducing interference between sites (Hill et al. 1966).

Examining variation in recombination rates among closely related species may provide clues as to the evolutionary forces affecting recombination, and therefore offer insight into the forces shaping the genome over time. Studies across virtually all eukaryotic kingdoms have determined that the distribution of recombination events is non-uniform. Indeed, there is variation in recombination rate across the genome; extreme rates are known as "hotspots" and "coldspots" in yeast and mammals, and reflect more of a quantitative change in other organisms. Hotspots are conventionally defined as a significant increase in recombination rate from the background recombination rate, usually ranging on orders of magnitude, that takes place in a small percentage of the genome. In humans this translates to about 80% of recombination taking place in less than 15% of the sequence (Myers et al. 2006). Although other organisms have not been assayed at the same resolution as yeast, mouse, and humans, many studies observe regions with a several fold increase above the background recombination rate. For example, in *Drosophila miranda*, several regions show recombination rates between 25-30 cM/Mb, several times the chromosomal average of ~5 cM/Mb (Cirulli et al. 2007; Stevison et al. 2010; McGaugh et al. 2012).

Recombination rates are variable between individuals, populations, and species; however, the causative factors underlying this variation are largely unknown. Through

the comparison of fine scale genetic maps of closely related species, it may be possible to identify features influencing fine scale and broad scale recombination patterns, as well as features that predict shifts in recombination landscapes between species. Such linkage map comparisons have the potential to answer questions such as how fast recombination rates change and how changes in recombination impact selection and evolution in natural populations. Theoretical work on the evolution of recombination and recombination modifiers has greatly contributed to these efforts, but this review will primarily focus on empirical work (For review of theoretical considerations, see (Charlesworth 1990; Barton 1995; Feldman et al. 1996; Lenormand et al. 2000; Martin et al. 2006; Barton 2010)).

The past few years have seen remarkable progress in the development of fine scale maps and in revealing novel modifiers of recombination rate. However, some comparisons of recombination maps show seemingly contradictory results, particularly in relation to conservation and divergence of recombination rates. Therefore, in this review, we discuss conservation and divergence in homologous meiotic recombination rate between closely related species. We begin by briefly considering sequence and epigenetic features known to affect recombination and the molecular evolutionary consequences of recombination rate variation. We proceed to discuss several issues surrounding the generation and analysis of recombination maps to understand why we would see conservation or divergence between some species and not others. Similarly,

we evaluate how constraints and regulating features could produce conflicting patterns of conservation and divergence when surveyed at different scales. Overall, we hope to highlight important questions pertaining to how recombination shapes genome evolution, and how studying closely related species can help answer them. Because the recombination literature is skewed towards humans, mice, and yeast, this review places an emphasis on these taxa, but attempts to incorporate other organisms whenever possible.

Determinants and Correlates

It is difficult to rigorously evaluate studies of recombination without considering various factors that may influence recombination rate. Control of recombination rate appears to be multifaceted, with molecular, environmental, and demographic factors all playing a role (for example, see (Wilfert et al. 2007), see also "Why would we see conservation of recombination rates between some species and not others?" below). Attempts to elucidate the determinants of recombination have identified numerous factors of influence, many of which appear to be conserved across eukaryotes (Lichten et al. 1995; Roeder 1997; Hassold et al. 2000; Keeney 2001; Petes 2001; Page et al. 2003). New hypotheses have emerged, for example, that variation in epigenetic features could explain variation in rates of recombination between closely related species (Myers et al. 2005; Ptak et al. 2005; Winckler et al. 2005; Myers et al. 2008). The continued study of

recombination between closely related species has the potential to illuminate more decisive determinants and how they change over time. Here, we focus on molecular patterns and features that appear to be associated with or impact recombination.

At the chromosome level, the prevailing pattern across organisms is that a minimum number of crossovers must be achieved for proper segregation. In humans, the number of crossovers is strongly correlated with the number of chromosomes, where one crossover per chromosome ensures proper segregation at meiosis (Fledel-Alon et al. 2009). In other organisms, crossover rates range from one crossover per chromosome arm to as many as five per chromosome (Beye et al. 2006). Animals with numerous shorter chromosomes, like the "microchromosomes" found in many birds, tend to have higher recombination rates, again, likely as a result of ensuring proper disjunction (Groenen et al. 2009). Along chromosomes in many organisms, rates of recombination tend to be higher toward the distal portions of the chromosome and low around the centromere, perhaps because repression of meiotic recombination by centromeric heterochromatin is also critical in proper segregation (Ellermeier et al. 2010).

At the genomic level, high recombination rates are positively and nearly ubiquitously associated with GC content, gene density, simple repeats, transposable elements, and a number of different sequence motifs (Thuriaux 1977; Gerton et al. 2000; Marais et al. 2001; Marais 2003; Jensen-Seaman et al. 2004; Meunier et al. 2004; Myers et al. 2005; Groenen et al. 2009; Wong et al. 2010). In particular, it appears that a 13-mer

degenerate motif may be responsible for recruiting recombination events in at least 40% of human hotspots (Myers et al. 2008).This motif binds the zinc finger protein PRDM9 in humans, and allelic variation controls hotspot activity in both humans and mice (Baudat et al. 2010; Berg et al. 2010). Relatedly, the Drosophila zinc finger protein, Trade Embargo (*trem*), initiates double strand breaks and is necessary for localization of the protein Mei-P22 to discrete foci on meiotic chromosomes, some or all of which are thought to mark sites for future double strand breaks (Lake et al. 2011). Thus, zinc finger proteins and sequence motifs may be major determinants of high recombination rate locations and recombination rate intensities at these locations.

Observations of divergent hotspot locations and usage among human individuals and between humans and chimpanzees has sparked a rigorous analysis of how epigenetics is involved in meiotic hotspot determination (Myers et al. 2005; Ptak et al. 2005; Winckler et al. 2005; Neumann et al. 2006). Subsequent studies show correlations between recombination hotspots and open chromatin, numerous histone modification patterns, and DNA methylation in yeast, mice, and humans (Berchowitz et al. 2009; Buard et al. 2009; Sigurdsson et al. 2009). Of particular note is the presence of a SET-methyltransferase domain in the *Prdm9* gene, which is responsible for the common chromatin feature trimethylation of lysine 4 of histone H3, or H3K4me3 (Baudat et al. 2010). H3K4me3 in yeast seems to be a prominent and pre-existing mark of active

recombination sites (Borde et al. 2009), potentially creating a link between sequence and epigenetic features affecting recombination.

Continued analysis between individuals and species will surely lead to a greater understanding of existing features and the discovery of novel ones. For example, the analysis of *Prdm9* across species has already produced fascinating results. Chimpanzee PRDM9 has dramatically different predicted binding sequence than human PRDM9 and seems to be the most divergent of all orthologous zinc finger proteins (Myers et al. 2010). Furthermore, *Prdm9* in other mammals shows rapid evolution with variation in zinc finger number and patterns of substitution suggestive of complex repeat shuffling (Oliver et al. 2009; Myers et al. 2010). Although there is no direct evidence that these changes have generated recombination rate differences, it is surely an intriguing area to be researched. If proven, this may provide an explanation as to how recombination hotspots are created and how they change over time (see section "Why would we see recombination rate conservation at some scales and not others?" below). Closely related species present the unique opportunity to study the evolution of features regulating and influencing recombination rate and should be central in future studies of this basic biological process.

Molecular Evolutionary Consequences of Recombination Rate Variation

Finally, recombination rate variation within and between closely related species allows evolutionary biologists to make conclusions as to whether selective or neutral forces are governing genomic landscapes. First postulated in a groundbreaking study, Begun and Aquadro (1992) found that recombination rate was positively correlated with nucleotide diversity in *Drosophila melanogaster*, but did not observe an association between recombination and *D. melanogaster-D. simulans* divergence . This pattern is interpreted to mean that natural selection, in particular selective sweeps and/or background selection, eliminates nucleotide variability in regions of low recombination (Smith et al. 1974; Charlesworth et al. 1993), and is supported by studies in several organisms (see **Table 1**). However, a similar association between recombination rate and nucleotide diversity may be predicted if recombination is mutagenic, but fewer studies have detected a correlation between recombination and nucleotide divergence between species and so have not met the prediction of the mutagenic hypothesis (but, see empirical studies (Brown et al. 1987; Brown et al. 1989; Strathern et al. 1995; Papavasiliou et al. 2000)).

Additionally, the interpretations are complicated by conflicting results in several organisms surveyed (Payseur et al. 2000; Baudry et al. 2001; Nachman 2002; Yi et al. 2004; Yi et al. 2005; Noor 2008; Tsai et al. 2010). Confounding factors that may lead to conflicting results are listed in **Table 1**, but there are several we would like to highlight

here. First, conflicting results may simply reflect taxon specific mutagenicity, but this hypothesis requires more empirical work. Second, correlations of diversity or divergence to recombination rate may change according to the scale with which recombination is assayed (Bussell et al. 2006; Spencer et al. 2006; Kulathinal et al. 2008; Noor 2008; Stevison et al. 2010), making it a priority to assess these measures using fine scale recombination over varying magnitudes. Third, and most relevant to the primary topic of this review, many studies up to this point have only assayed recombination in one species of interest, assuming recombination rates are conserved. It remains unresolved as to whether the selection or mutagenic hypothesis primarily accounts for the observed pattern, but perhaps with increasing amounts of recombination and sequence data, we will be able to make firmer conclusions.

Conservation and Divergence of Recombination

Following the progressive discoveries in diverse species that recombination events are non-random across the genome, one of the most exciting and surprising findings has been the realization that recombination rates sometimes change, even within species or between closely related species. The fact that recombination rates are variable and heritable implies that recombination itself can evolve in response to natural selection (Chinnici 1971; Charlesworth et al. 1985; Otto et al. 1998). Furthermore, evidence from human recombination hotspots seems to show that this change can occur

quickly on an evolutionary timescale, with hotspots emerging and disappearing in as little as 120,000 years and certainly within the six million years human divergence from chimpanzee (Ptak et al. 2005; Winckler et al. 2005; Jeffreys et al. 2009). However, conservation of recombination between closely related species has also been detected at varying scales (see **Table 2**), raising many questions. (1) How does the methodology by which recombination is measured affect estimates of recombination rate? (2) Why would we see conservation of recombination rates between some species and not others? (3) Why would we see conservation at some scales and not others? (4) Finally, and perhaps most fundamentally, should we expect to see conservation between closely related species? Here we comprehensively review empirical studies that compare recombination rates between closely related species, and speculate on the answers to these questions.

How does the methodology by which recombination is measured affect estimates of recombination rate?

The construction of a recombination map can dramatically affect the estimate of recombination rate depending on the methodology employed (and associated biases). Three methods are commonly used for estimating recombination rate: Linkage Disequilibrium (LD) mapping, sperm-typing, and direct mapping using polymorphic markers (see **Table 2** for examples of recombination maps made with these approaches, **Table 3** for potential strengths and weaknesses). The first two methods are used

primarily with human data (but see (Guillon et al. 2002; Ptak et al. 2004; Kim et al. 2007; Arguello et al. 2010)), with the labor and resource intensive direct mapping applied more in other model organisms. Fundamentally, the major differences between these measures are (1) whether the recombination rates measured are current vs. historical, and (2) whether the recombination rates measured reflect a population average or focus on a particular individual or set of individuals.

There has been some doubt as to whether LD consistently and accurately predicts hotspots (Jeffreys et al. 2005; Reed et al. 2006). For instance, Coop *et al.* (2008) estimated that 40% of crossovers occurred outside of LD-predicted hotspots, but Khil *et al.* (2010) suspect this may be an overestimate due to the way hotspots were measured using particular populations. Khil *et al.* also found that 26-32% of crossovers happened outside of European population LD-predicted hotspots, however this discrepancy disappeared when hotspot locations of other populations were taken into account. There are also specific examples of LD-predicted hotspots being absent when checked with sperm typing (Kauppi et al. 2005), although such inconsistencies appear to be the exception rather than the rule (Jeffreys et al. 2000; Jeffreys et al. 2001; Yauk et al. 2003). In pedigrees and controlled crosses, increasing sample size and marker coverage changes the way we measure and perceive recombination, similar to the "Beavis effect" for mapping (Beavis et al. 1994; Beavis 1996). For example, the original honeybee linkage map used 94-142 individuals and 365 RAPD markers for a total map length of 3450 cM

(Hunt et al. 1995); the newer map used 541 markers and 283 individuals for a total length of 4061.2 cM (Solignac et al. 2004). With more accurate technologies, the honeybee genome increased in length from 178 Mb in 1974 (Jordan et al. 1974) to 262 Mb in 2006 (The Honeybee Genome Consortium), thereby decreasing the average recombination rate from 19.38 cM/Mb to 16 cM/Mb today. Similarly, in the chicken, the current map used 9268 markers for a total length of 3228 cM (Groenen et al. 2009), substantially smaller than the 4200 cM previously estimated with 2261 markers (Schmid et al. 2005). However, obtaining enough markers to detect fine scale recombination is resource intensive, thereby producing maps that range in scale from kilobases (hereafter referred to as "fine scale") to hundreds of kilobases (hereafter referred to as "intermediate scale") to tens of megabases (hereafter referred to as "broad scale") to whole genomes. This is relevant, as recombination rate conservation and divergence between species is scale dependent (see below).

Other indirect quantitative approaches also exist, such as immunostaining as used by Dumont and Payseur (2011) in Murid rodents and Double Strand Break (DSB) mapping, most commonly used in yeast (Gerton et al. 2000; Buhler et al. 2007; Mancera et al. 2008; Dumont et al. 2011). Of course, choosing an approach is constrained by the organism and resources available, and researchers must be aware of limitations when making generalizations and conclusions.

Why would we see conservation of recombination rates between some species and not others?

There are several pertinent issues to consider when comparing recombination rates between closely related species. First, differential action of selection, or selection in changing environments, could give rise to differences between species (Chinnici 1971; Charlesworth et al. 1985; True et al. 1996). For example, artificial selection may have increased recombination rates in the domesticated species of chicken, honeybee, and many plants (Rees et al. 1974; Burt et al. 1987; Otto et al. 2001; Ross-Ibarra 2004; Wilfert et al. 2007; Groenen et al. 2009) in comparison to their wild progenitors.

Second, imprecise measures of physical distance can lead to inferred differences in recombination rate. Large insertions, deletions, or inversions can affect the recombination fraction between two points in the genome because single crossover events may often lead to aneuploidy. Even if such changes do not change the actual recombination "rate" (e.g., centiMorgans per megabase), if a large insertion is present in one taxon but not another, it may appear that recombination rates have diverged between species, when in fact recombination is being measured over intervals of unequal size. Moreover, chromosomal rearrangements, such as inversions, reduce the observed amount of recombination in heterokaryotypes (Hartl and Jones 2004; pp. 319-324). This repression of recombination associated with inversions extends outside the inversion several megabases (Mb), thereby producing broader scale changes in recombination rate (Kulathinal et al. 2009). Furthermore, inversion heterozygotes exhibit

increased recombination further outside the inverted region, known as the "interchromosomal effect" (Schultz et al. 1951). Large insertions and deletions have the potential to produce comparable outcomes to inversions. In comparing the same intervals between two closely related species in the absence of a genome sequence, one runs the risk of concluding increased divergence between species when in actuality, an inversion, insertion, or deletion segregating in only one species is obscuring their comparable recombination rates.

Third, recombination is variable within individuals and populations (Brooks et al. 1986; True et al. 1996; Carrington et al. 2004; Neumann et al. 2006; Graffelman et al. 2007; Coop et al. 2008; Paigen et al. 2008; Cheng et al. 2009; Dumont et al. 2009; Kong et al. 2010). This variation may stem from actual heritable variation in recombination rates among individuals, variation within an individual among regions of its genome (as discussed above), or from environmental variation. Because of potentially dramatic within-population variation, one must use caution in making generalizations about species as a whole or presumptions that a single linkage map describes the species. Without first surveying recombination within the populations, subspecies, or species of interest, it will be impossible to conclude true conservation or divergence between groups. These potential challenges in estimating recombination rate can translate into real or perceived differences between closely related species.

Why would we see recombination rate conservation at some scales and not others?

Maps produced at varying scales provide an unexpected and seemingly contradictory view of conservation and divergence between species (see **Table 2**). In mice and humans, conservation of recombination rate was identified at a broad scale, but divergence of recombination hotspots was identified at a fine scale, suggesting that there may be differential regulation among the scales (Myers et al. 2005; Coop et al. 2007), or perhaps, simply that we do not understand the relationship between fine scale and broad scale recombination patterns.

Divergence of fine scale recombination rates observed in yeast and humans is likely caused by the rapid turnover of hotspots due to biased gene conversion and meiotic drive. In the process of double stranded break repair, one pathway leads to crossing over and exchanging of content between homologous chromosomes; the other pathway leads to gene conversion (Boulton et al. 1997; Marais 2003). Biased gene conversion favors one allele over the other, in which case the initiating hotspot is replaced by a copy of its homolog, effectively suppressing subsequent recombination. Simulations and empirical evidence have demonstrated the self-destructive nature of hotspots through overtransmission of recombination suppressing alleles (meiotic drive), creating what is known as "the hotspot paradox" (Boulton et al. 1997; Jeffreys et al. 2002).

Recent evidence might provide the elusive answer as to how new hotspots are created and regulated to counteract losses due to the hotspot paradox. The zinc finger protein PRDM9, confirmed to play a significant role in recombination in human and mouse, contains a zinc finger encoding region with a minisatellite structure (Baudat et al. 2010). This particular structure may confer a strong potential to generate variability by recombination or replication slippage within the array. Indeed, studies documented variability in the contact residues predicting DNA binding between human populations (Baudat et al. 2010; Berg et al. 2010; Parvanov et al. 2010), and the number of zinc fingers and their contact residues vary significantly across rodents, primates, and other Metazoans (Oliver et al. 2009). The changing of contact residues could create a new family of hotspots by the binding of the protein to new sequence motifs, thereby counteracting the loss of hotspots due to biased gene conversion. Regardless of the mechanism, it is clear by the constant flux of hotspots that the precise locations seem to be unconstrained, allowing for divergence at the hotspot level.

It is believed that broad scale recombination is controlled in a different, but potentially non-mutually exclusive manner. Above all, the broad scale rate is defined by the necessity of one crossover per chromosome to ensure proper disjunction (Hassold et al. 2004; Fledel-Alon et al. 2009). Other chromosomal properties, including size and number, are correlated with this trend. This selective constraint is likely a large influence on the conservation of broad scale rates between closely related species. At the

intermediate scale, conclusions are more vague. Regional recombination is decidedly influenced by crossover interference, where a crossover in one location prevents another crossover from occurring close by (Foss et al. 1993; Hillers 2004; Stahl et al. 2004; Copenhaver 2005). Regional properties of chromosomes have an impact as well, clearly shown by the lack of crossovers in the centromeric region and typically high number of crossovers near the telomeres. Overall, perhaps these broader scale processes are more likely to be conserved, and hence preserve recombination at this scale as a byproduct.

The connection between the different scales is also unclear. In humans, there are an estimated 60-80,000 hotspots across the genome (Khil et al. 2010), occurring in clusters every 60-90 kb with individual hotspots separated by 1-7 kb within each cluster (Jeffreys et al. 2001; Kelmenson et al. 2005) and no region greater than 200kb in which recombination is absent (Myers et al. 2006). It appears that non-hotspot recombination is relatively rare, with sperm genotyping studies showing very low levels of background, non-hotspot recombination, most likely in weaker and polymorphic hotspots (Jeffreys et al. 1998; Jeffreys et al. 2001; Jeffreys et al. 2002; Jeffreys et al. 2005; Jeffreys et al. 2005). It is proposed that larger scale variation may be a product of the varying density or intensity of these hotspots in different regions of the genome (Nachman 2002; Myers et al. 2006).

There are several hypotheses as to which selective pressures influence the overall process and distribution. Hey (2004) proposed that individual hotspots may arise as a

byproduct of linkage disequilibrium between genes that are being selected (Hey 2004). If selection favors a recombinant haplotype, then it will favor chromosomes with high recombination between the two genes, thereby producing variation in location and intensity of local recombination rates that fluctuate depending on the genes under selection and the linkage disequilibrium patterns. Kauppi *et al.* (2004) put forward three hypotheses; one, that the distribution of recombination is governed by a balance between the need for recombination to ensure proper segregation during meiosis and the need to minimize the breakdown of favorable haplotypes. Two and three, more mechanistically, that restrictions on crossover position potentially facilitate the optimal mechanical/biochemical function of chiasmata in chromosome segregation or restrictions on position confine double strand breaks to regions that are most conducive to efficient assembly of machinery and repair.

The answer is plausibly a combination of all of these propositions. The continued analysis of recombination rates assayed at different scales, combined with theory that could test selective constraints at these scales, will help to determine why we would see conservation in recombination rate at some scales and not others.

Should we expect to see conservation between closely related species?

Just as we should expect to see conservation in sequence between closely related species, we should expect to see conservation in recombination (Dumont et al. 2008).

Presumably, with sequence similarity comes, on average, greater sharing of features that influence recombination. Divergence should be seen as a departure from the null hypothesis, but only if there is a significant increase in divergence relative to variability of recombination within species, something rarely studied.

Organisms with resolution of recombination rate at the fine scale (e.g. human, mouse, and yeast) show that hotspots do indeed diverge over time. Other organisms with maps at an intermediate to broad scale generally portray a trend of conservation (see **Table 2**). From the evidence we have, we can determine that hotspots are not conserved due to their transient nature. Over long periods of time, enough changes occur on the fine scale to produce broad scale change. Additionally, on a broad scale, shared constraints will hold the rate more constant, so that rates are more conserved between closely related species. These processes are compounded by changes in the genome such as inversions, translocations, insertions, and deletions that lead to altered recombination rates over time.

Future studies should strive to confirm both these trends, and the mechanisms behind them, to better understand the impact recombination has on the genome, and which features of the genome regulate recombination. A combination of theoretical and empirical work will be necessary. Theoretically, models and simulations can inform role of selective and neutral processes and different population parameters that could lead to conservation or divergence of recombination rates. Empirically, in chimpanzees,

comparisons of human-chimpanzee recombination rates should be extended beyond isolated regions of the genome. Recombination should also be surveyed in a broader set of species, with a particular attempt to obtain a resolution of less than 10 kb in order to observe the presence or absence of hotspots, albeit this is a major challenge in non-model systems. The ability to make cross species generalizations about recombination rates will have an impact on our understanding of genome evolution, therefore implicating diverse topics such as human health, selection and neutrality in the genome, and speciation and mapping studies.

Conclusions

Several patterns emerge in reviewing data from these recombination maps. First, the labels such as conservation and divergence are somewhat misleading. No pair of species studied to date exhibits complete divergence or conservation of all studied hotspots or regions of high recombination, and furthermore, many studies report only a few intervals across the genome that can be categorized in these terms. Therefore, it is necessary to attempt to incorporate a standard, or at least explicitly defined, set of parameters when discussing divergence. Of course, the vast range of scales at which recombination has been, and continues to be, surveyed makes this difficult. Because recombination is considered a quantitative genetic trait displaying variation and heritability that can be acted upon by natural selection, it shares features with other

phenotypes and can be discussed in similar terms. A standard approach is to discuss sequence divergence as a percentage, and here we recommend recombination rate divergence to be treated in a similar manner. It is appropriate to express intervals conserved or diverged out of the total number surveyed, also citing the percentage of the total of the physical and recombinational maps that was surveyed.

With more genome sequences becoming available, the decreasing cost of genotyping, and sophisticated software and technology at our fingertips, detection of recombination can be completed with greater feasibility. These projects should be undertaken with several parameters in mind. First, recombination variation should be assayed within and across populations within species before making conclusions about differences between species. This is necessary to make conclusions regarding conservation or divergence between species, but will also provide a perspective on the speed at which recombination rate is changing. Second, genome sequences should be available for the populations or species involved, and particularly for the strains being surveyed. This data will ensure that no chromosomal rearrangements or insertions/deletions exist, and will facilitate analysis of features such as motifs and measures of diversity and divergence. Surveying diversity and divergence will also provide a better understanding of forces at work in the genome. Third, examining recombination at multiple scales will aid in the interpretation of different constraints influencing the distribution of recombination events, especially when compared

between populations and species. Finally, in interpreting results, population history and the methodology used to infer recombination should be taken into account to avoid biases and complications.

Fine scale recombination maps can help us address essential questions, such as how variation is created and maintained within and between species. Recombination shapes features of the genome and creates new allelic combinations that allow increased adaptability in all sexual organisms. However, many researchers have assumed that recombination rate is invariable among individuals and between species, which we now know is inherently false. Indeed, knowledge of fine scale variation in crossover rate is essential in modeling genome evolution, population genetics studies, genome wide association studies, and inferring evolutionary processes. Thus, results indicating how recombination rate is distributed in the genome will have implications in human health, molecular evolution, and the way we study genetics. The role of recombination in genome evolution is a fundamental issue in understanding basic biological processes, and while much progress has been made, many questions remain to be answered.

Table 1: Relationships between recombination rate and measures of diversity and divergence

Species 1	Species 2	Time since divergence (My)	Relationship between Diversity and Recombination	Relationship between Divergence and Recombination	Confounding factors	Claims Support for Natural Selection or Mutagenic Hypothesis?	Source
Human	Chimpanzee	5-6	Positive, $p = 0.06$, $r = 0.678$	Positive, $p = 0.006$, $r = 0.866$	Only surveyed PAR1; recombination rate only surveyed in humans	Claims support for mutagenic hypothesis for recombination	(Bussell et al. 2006)
Human	Chimpanzee	5-6	Positive, $p = 0.01$, $r = 0.127$ (Stephens 2001), $p = 0.037$, $r = 0.257$ (SeattleSNPs), $p = 0.032$, $r = 0.249$ (NIEHS SNPs)	Positive, $p = 0.037$, $r = 0.259$ (Seattle), $p = 10^{-6}$, $r = 0.214$ (chimp shotgun), $p = 4x10^{-4}$, $r = 0.289$ (chimp BAC), $p = 0.204$, $r = 0.264$ (Baboon BAC)	Recombination rate only surveyed in humans	Claims support for mutagenic hypothesis for recombination	(Hellmann et al. 2003)
Human	Chimpanzee	5-6	Positive, $p<0.001$, $r = 0.735$	Uncorrelated, $p = 0.57$, $r = 0.141$	Used recombination estimates from numerous studies, which used various methods to estimate recombination and may be imprecise; used SNPs in close proximity to exons, which may be influenced by other forces; recombination rate only surveyed in humans	Claims support for natural selection hypothesis for recombination	(Nachman 2001)
Human	Chimpanzee	5-6	Positive, $p = 0.041$,	Uncorrelated, $p = 0.726$,	Recombination rate only surveyed in humans;	Claims support for natural selection	(Nachman et al. 1998)

			r = 0.775	r = 0.164	based on small sample size (seven data points)	hypothesis for recombination	
Mus musculus	*Mus spicilegus*	3	Uncorrelated, p = 0.83 (within subpopulations, p = 0.045), r = 0.06 (within subpopulations, positive, r = 0.46)	Uncorrelated, p = 0.65, r = 0.11	Recombination estimates only from *M. m. domesticus*; used sequence data near functional genes; small sample size	Claims support for natural selection hypothesis for recombination	(Takahashi et al. 2004)
Mus mus domesticus	*Mus caroli*	2.5	Positive, p = 0.0972 r = 0.902	Negative, p = 0.5663 r = 0.434	Imprecise estimates of recombination; small sample size; only surveyed at four loci	Inconclusive based on available data	(Nachman 1997)
Sacchromyces cerevisiae	*Sacchromyces paradoxus*	5-20	NA, Uncorrelated	NA, Uncorrelated	Low frequency of sex and outcrossing	Inconclusive; suspect mutagenic effect of recombination swamped by the mutations occurring in intervening asexual generations	(Tsai et al. 2010)
Drosophila melanogaster	*Drosophila simulans*	2.5-3.4	Positive, p = 0.0007	Uncorrelated, p = 0.983	Recombination only surveyed in *D. melanogaster*; low quality estimates of recombination; only surveyed at 20 gene regions	Claims support for natural selection hypothesis for recombination	(Begun et al. 1992)
Drosophila simulans	*Drosophila melanogaster*	2.5-3.4	Positive, $p = 8.5 \times 10^{-8}$	NA, p = 0.03	Recombination only surveyed in *D. melanogaster*; low quality map for *D. melanogaster;* only surveyed X	Claims support for natural selection hypothesis for recombination	(Begun et al. 2007)

					chromosome		
Drosophila pseudoobscura	*Drosophila persimilis*	0.5-1	Positive, **≤500 kb window**: Intronic p = 0.0156, r = 0.488, Intergenic p = 0.0020, r = 0.588 **2 Mb window**: Intronic p = 0.8322, r = 0.045, Intergenic p = 0.3260, r = 0.205	Positive, **≤500 kb window:** Intergenic p = 0.0006, r = 0.635 **2 Mb window:** Intergenic p = 0.0013, r = 0.607	Shared ancestral polymorphism; hybridization; Recombination only surveyed in *D. pseudoobscura*	Claims support for mutagenic hypothesis for recombination	(Kulathinal *et al*. 2008)
Drosophila pseudoobscura	*Drosophila miranda*	2-3	Intronic p = 0.004, r = 0.333, Intergenic p = 0.258, r = 0.133	Intronic p = 0.128, r = 0.180, Intergenic p = 0.559, r = 0.069	Recombination only surveyed in *D. pseudoobscura*	Claims support for natural selection hypothesis for recombination	(Stevison and Noor 2010)
Beta vulgaris maritima (sea beet, wild relative of the sugar beet)	*Beta macrocarpa*	NA	Positive, p = 0.007, r = 0.226 (based on number of distinct alleles) p = 0.084, r = 0.117 (based on heterozygosity) *calculated using Θ	Negative, Not significant, r = -0.02	Recombination estimates for *Beta vulgaris maritima* come from *B. vulgaris vulgaris*, which may not be an accurate assumption; recombination estimates are coarse	Claims support for natural selection hypothesis for recombination	(Kraft et al. 1998)
Lycopersicon peruvianum (tomato)	*S. ochranthum* or *S. lycopersicoides*	5.8-18.6	Positive, p = 0.089, r = 0.471 *calculated using Θ	Uncorrelated, p = 0.640, r = 0.159	Demographic processes and life history traits may confound diversity and recombination estimates; genomic recombination rate used is based on a map from *L. esculentum* x *L. pennellii*, which may not be an accurate assumption	Demographic processes have a strong influence on shaping patterns in the genome in combination with mutation rate and/or selective constraint	(Roselius et al. 2005) see also (Stephan et al. 1998)
Lycopersicon chilense			Uncorrelated, p = 0.498, r = 0.198 *calculated using Θ	Uncorrelated, p = 0.852, r = 0.064			
Lycopersicon			Positive,	Uncorrelated,			

hirsutum			p = 0.052, r = 0.528 *calculated using Θ	p = 0.915, r = 0.036			
Zea mays mays (corn)	*Zea mays parviglumis*	7,500-9,000 years	Positive, p = 0.007, r = 0.65 (using $4N_{Chud87}$); Uncorrelated when recombination measured with physical measure *R* or $4N_{Chud01}$ *calculated using Θ	Uncorrelated, NA	Numerous studies produce different results; different recombination estimates produce different results; low power; complicated by demography	Inconclusive; Not accounted for by Demography model, selection plays a role	(Tenaillon et al. 2001; Tenaillon et al. 2002; Tenaillon et al. 2004)

Table 2: Relationships between recombination rate and measures of diversity and divergence

Species 1	Species 2	Time since divergence (My); Approximate Sequence Divergence (%)	Region of genome surveyed	Approximate percentage of physical genome surveyed in between species comparison (%)	Methodology used to measure recombination rate	Number of markers	Marker distance	Size of intervals compared between species	Sample size	Conservation or divergence?	Source
Human	Chimpanzee	5-6; 1	Several known hotspots; 3 500kb regions located on 4q26, 7q21, 7q31	<1	Linkage disequilibrium (LD)	Hotspot specific	Hotspot specific	Hotspot specific	Hotspot specific	15/18 hotspots are divergent; Divergent in 10 kb windows; Divergent in 14 additional 160 kb windows	(Winckler *et al.* 2005)
Human	Chimpanzee	5-6; 1	2 regions totaling 14 Mb on Chr. 21	<1	Linkage disequilibrium (LD)	30,611 (11,642)	440 bp (1.2 kb)	50 kb	8 (71)	36/39 hotspots are divergent; Significantly, but weakly correlated at 50 kb (r = 0.276); Total recombination rates significantly, but weakly correlated (r =	(Ptak *et al.* 2005)

										0.216)	
C57BL/6J x CAST/EiJ	8 Heterogeneous stock mice (A/J, AKR/J, BALB/cJ, DBA/2J, C57BL/6J, LP/J, I, and RIIIS/J)	0.2375-0.475; <1	Chr. 1	7	SNP genotyping of controlled crosses	1059 (873)	225 kb (225 kb)	550 kb	6028 (2293)	Significant, but weak regional correlation (r = .38)	(Paigen et al. 2008) recombination data for HS mice from (Shifman et al. 2006)
Mus musculus domesticus x Mus musculus musculus	Mus musculus domesticus x Mus musculus castaneus	0.2375-0.475; <1	Genome wide	84.8 (84.0)	SNP genotyping of controlled crosses	186 (197)	1.2-65.6 Mb (mean 15.8 Mb)	1.2-65.6 Mb (mean 15.8 Mb)	580 (554)	31/131 intervals divergent	(Dumont *et al*. 2010)
Gallus gallus (chicken)	*Taeniopygia guttata* (zebra finch)	100; NA	Genome wide	92	Pedigree SNP genotyping	1404 (9268)	800 kb (100kb)	1 Mb	1079 (235)	Significantly correlated when comparing 275 intervals (r = 0.5)	(Backstrom et al. 2010) (recombination data for chicken from Groenen *et al.* 2009)
Gallus gallus (chicken)	*Acrocephalus arundinaceus* (great reed warbler)	80-100; NA	Regions of 9 chromosomes (ranging from 5%-66% coverage)	26% of chicken genome	Pedigree SNP genotyping	46; (NA)	NA	NA	812; (NA)	Great reed warbler has substantially shorter map than chicken	(Dawson et al. 2007) (recombination data for chicken

											from http://www.ncbi.nlm.nih.gov/genome/guide/chicken/)
Drosophila pseudoobscura	*Drosophila persimilis*	0.5-1; NA	Chr. 2	23	SNP genotyping of controlled crosses	130 (50)	240kb (avg. 466 kb)	500 kb	1440 (1294)	Divergent in 5/38 intervals	(Stevison and Noor 2010) (recombination data for *D. pseudoobscura* from Kulathinal *et al.* 2008)
Apis mellifera (honey bee)	*Apis florea*	8-10; NA	Chr. 3 and Chr. 12	12.9	SNP genotyping	Chr. 3: 8, Chr. 12: 10 (2000 genome wide)	Chr. 3: 1.2 Mb, Chr. 12: 780kb (100 kb)	Chr. 3: 1.2 Mb, Chr. 12: 780kb	120 (92-187)	Conservation in 19/19 intervals	(Meznar et al. 2010) (recombination data for *A. mellifera* from Solignac *et al.* 2007)
Sacchromyces cerevisiae	*Sacchromyces paradoxus*	5-20; 13	Chr. 3	3	Linkage disequilibrium (LD)	464 European, 232 Far East	600 bp European, 1200 bp Far	Hotspot specific	20 strains (51 meioses	6/10 hotspots are divergent	(Tsai *et al.* 2010) (recombination

						(52,000 genome wide)	East (78 bp)		; 204 spores)		data for *S. cerevisiae* from Buhler *et al*. 2007, Mancera *et al*. 2008)

Table 3: A comparison of recombination rates between species pairs

Approach	Linkage Disequilibrium	Sperm-typing	Pedigrees and crosses
Description	LD-based genetic maps use statistics to estimate historical recombination indirectly from patterns of allelic associations in samples from natural populations	Individual or pooled sperm is analyzed for linkage disequilibrium blocks using allele-specific PCR directed to heterozygous SNP sites	Genotype markers in pedigrees or controlled crosses
Current vs. historical recombination	Historical	Current	Current
Application	Used in humans, chimpanzees; Can be applied to other organisms depending on population history	Used in humans, mice	Used in mammals, birds, yeast, plants, insects
Potential Strengths	Relatively quick and inexpensive way to assay whole genome	Direct assay of current recombination; Looking at variation within and between individuals; Looking at specific hotspots; Detecting gene conversion events	Direct assay of current recombination; Can apply across whole genome; Can get sex-specific rates for both sexes;
Potential Weaknesses	Breaks in LD are not always hotspots; Cannot necessarily detect recent recombination events; Sex-averaged; Made with heterogeneous populations; Known to be influenced by	Male specific (although this can also be a strength as it gives sex-specific information that LD cannot); Labor intensive; Can't assay whole genome or population	Sensitive to sample size, number of markers, and unknown variation in genome size, structure, and individual variation in recombination rate;

	genetic drift, demographic factors, natural selection, variable mutation rates, and gene conversion		Difficult to achieve fine scale; Labor and resource intensive
Further reading	Ardlie *et al.* 2002, Clark *et al*. 2010, Slatkin 2008	Carrington and Cullen 2004, Jeffreys *et al.* 2001, Jeffreys and Neumann 2002	Kong *et al.* 2010, see studies in **Table 2**

2. Recombination modulates how selection affects linked sites in *Drosophila*

Suzanne E. McGaugh, **Caiti S. S. Heil**, Brenda Manzano-Winkler, Laurence Loewe, Steve Goldstein, Tiffany L. Himmel, Mohamed A. F. Noor

Citation: McGaugh, S. E., C. S. Heil, et al. (2012). "Recombination modulates how selection affects linked sites in Drosophila." PLoS Biol 10(11): e1001422.

Smukowski Heil Contribution

I was responsible for all *Drosophila miranda* data in this manuscript, which includes conducting the experimental cross, designing SNP markers, and analyzing all genotypic results. I collaborated with a statistician, Tiffany Himmel, to statistically analyze conservation and divergence of recombination rates, and did the analyses for the global recombination modifier in *D. miranda*. I furthermore looked at various genomic correlates with recombination. Finally, I assisted in the writing and editing of the manuscript.

Introduction

Homologous meiotic recombination has an important role in molecular evolution. Sufficient recombination uncouples the evolution of different sites on the same chromosome allowing positive or negative selection at one site to act independently from selection at another site. If there is less than effectively free recombination between two selected sites, then linkage results in selection at one site interfering with selection at another site. This has been termed “Hill–Roberson

interference" (Hill et al. 1966; McVean et al. 2000; Comeron et al. 2002; Charlesworth et al. 2009; Loewe et al. 2010; Charlesworth 2012). Hill–Robertson interference increases the probability of fixation of deleterious mutations, decreases the probability of fixation of advantageous mutations, and reduces overall DNA sequence diversity. Thus, the breakdown of linkage disequilibrium between loci experiencing Hill–Robertson interference allows selection to act more efficiently, purging deleterious mutations and accelerating adaptation (Hill et al. 1966; McVean et al. 2000; Comeron et al. 2002; Charlesworth et al. 2009; Loewe et al. 2010; Charlesworth 2012).

Such indirect effects of recombination on the genome (Webster et al. 2012) result in a positive association between the rate of recombination and adaptive evolution (Colegrave 2002; Bourguet et al. 2003; Goddard et al. 2005). For example, recombination rate is positively associated with codon usage bias, whereby those codons coded by the most abundant tRNAs are "preferred" and used more often (Hey et al. 2002; Haddrill et al. 2007). Recombination has direct effects on a genome sequence as well, because recombination influences base composition through biased gene conversion and the distribution of repetitive elements, hotspot sequences, and indels (Begun et al. 1992; Charlesworth et al. 1994; Comeron et al. 1999; Duret et al. 2008; Myers et al. 2010; Webster et al. 2012). Understanding the magnitude of indirect effects in light of these direct effects has proved challenging (Webster et al. 2012).

One striking association is a positive relationship of local recombination rate and nucleotide diversity (Aguade et al. 1989; Stephan et al. 1989; Begun et al. 1992). Originally described in *Drosophila melanogaster* (Begun et al. 1992), the positive relationship between recombination rate and nucleotide diversity has been demonstrated in a wide range of taxa, including humans, mice, yeast, maize, and tomatoes (reviewed in (Smukowski et al. 2011)). It is not fully understood how much of this relationship results from recombination's indirect versus direct effects on the genome. For instance, mutations created during crossing over or double-strand break repair may generate new polymorphisms and hence increase diversity (Esposito et al. 1993; Strathern et al. 1995; Rattray et al. 2001; Lercher et al. 2002; Filatov et al. 2003; Kulathinal et al. 2008; Hicks et al. 2010). Alternatively, recombination may indirectly influence genetic diversity by mitigating the genomic footprint of selective sweeps and background selection (Smith et al. 1974; Charlesworth et al. 1993; Nordborg et al. 1996).

One way to distinguish between these general explanations is to evaluate the relationship of between-species nucleotide divergence at neutral sites and local recombination rate, because truly neutral mutations are substituted at the same average rate between species as they appear between generations, even if linked to sites under selection (Kimura 1983; Birky et al. 1988). This allows us to predict that both within-species nucleotide diversity and between-species nucleotide divergence would have a positive relationship with local recombination rate (Begun et al. 1992), if the

recombination–diversity association was purely caused by mutation. In contrast, selective sweeps and background selection will cause an association between recombination and within-species nucleotide diversity, but not a relationship between recombination and between-species nucleotide divergence (Smith et al. 1974; Birky et al. 1988). The absence of an association of between-species nucleotide divergence and local recombination rate suggests that variation in recombination rate translates to variation in the efficiency of selection (Begun et al. 1992). Past work relating nucleotide divergence to recombination rate found no relationship between these two variables in several species of *Drosophila*, mouse, beet, yeast, and other species (Begun et al. 1992; Kraft et al. 1998; Takahashi et al. 2004; Noor 2008; Stevison et al. 2010; Smukowski et al. 2011; Mackay et al. 2012). Furthermore, in several species, evidence indicates that segregating ancestral polymorphisms may be responsible for correlations between divergence and recombination rate ((Begun et al. 2007; Cutter et al. 2010; Lohmueller et al. 2011), also suggested by (Kulathinal et al. 2008; Noor 2008)).

The test above, however, implicitly assumes that local recombination rates are conserved between the two species used to generate the nucleotide divergence measure. If recombination rate has diverged between the two species, no relationship between local recombination rate and nucleotide divergence may be detected even when recombination is mutagenic (see **Figure 1**). Recombination rates, especially at fine scales, are often not conserved among closely related species, as is the case between humans

and chimpanzees (Ptak et al. 2004; Ptak et al. 2005; Winckler et al. 2005); thus, the assumption of conservation of recombination rates may be violated in previous studies, and a more definitive understanding of the diversity–recombination association awaits estimates that are free from this assumption.

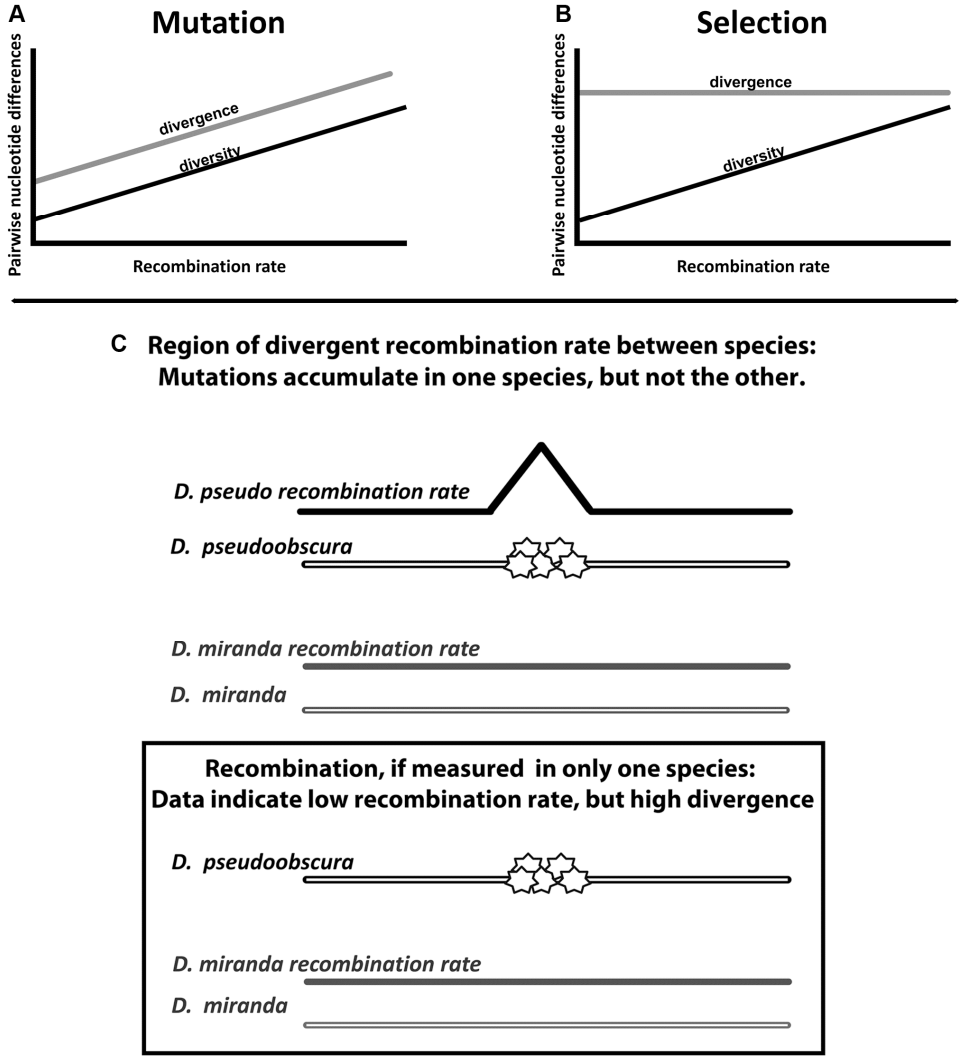

Figure 1: Expectation of the relationship between divergence and recombination rate if the recombination–diversity positive correlation is the result of recombination being mostly mutagenic or the result of recombination's effect on selection at linked sites.

Though there are theoretical expectations concerning how recombination rate should affect selection efficiency (Kim 2004; Loewe et al. 2007), it is unclear empirically whether variation in local recombination rates translates into significant variation in the efficiency of selection (Webster et al. 2012). Several empirical studies have tackled this problem (Pal et al. 2001; Betancourt et al. 2002; Presgraves 2005; Haddrill et al. 2007; Bullaughey et al. 2008; Larracuente et al. 2008; Weber et al. 2009; Cutter et al. 2010), and many findings suggest that recombination rate influences the efficiency of positive or negative selection in regions of moderate or high recombination. Still, various confounding factors (e.g., biased gene conversion, gene density) may produce spurious correlations between both recombination and substitution rate, and some authors suggest that there is no strong empirical evidence for recombination affecting the efficiency of selection (apart from reduced selection in regions with essentially no recombination).

The *Drosophila pseudoobscura* system is ideal for pursuing questions about recombination rate variation and its molecular evolutionary consequences. The average crossover rate of *D. pseudoobscura* (about 7 cM/Mb in females) is over twice that of *D. melanogaster* (Ortiz-Barrientos et al. 2006). There is also considerable fine-scale (<200 kb windows) variation in the local recombination rate within the genome of *D. pseudoobscura* and within the genome of its sister species, *D. persimilis* (Cirulli et al. 2007; Kulathinal et al. 2008; Stevison et al. 2010). While some recombination data are available

for *D. pseudoobscura* and *D. persimilis*, these sister taxa interbreed in the wild (Dobzhansky 1973; Powell 1983; Machado et al. 2002) and are, therefore, not ideal for examining the divergence–recombination association. For example, shared polymorphism due to hybridization and recent speciation may be responsible for the positive divergence–recombination association found in a previous study (Kulathinal et al. 2008) see also (Cutter et al. 2010; Lohmueller et al. 2011). Fortunately, a third species exists (*D. miranda*) that is phylogenetically close to *D. pseudoobscura* but does not interbreed with *D. pseudoobscura*. Since there is still some residual shared ancestral polymorphism (Nowell et al. 2011), we also obtained the genome sequence for a slightly more distantly related outgroup species, *D. lowei* (**Figure 2**). Sequence from *D. lowei* is useful for generating a proxy for neutral mutation rate across the genome.

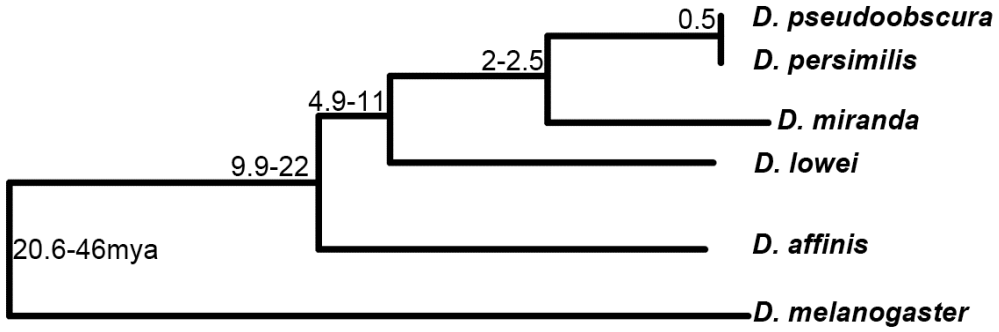

Figure 2: Relationships of study species

In this work, we generate and compare two fine-scale recombination maps for *D. pseudoobscura*, which each cover approximately 43% of the *D. pseudoobscura* physical genome and one fine-scale recombination map that covers approximately 31% of the *D.*

miranda physical genome. In order to circumvent the assumption of classic studies, we analyze the relationship of local recombination rate to nucleotide diversity and divergence in regions with very similar recombination rates between the two species. By employing a linear model framework to account for multiple covariates, we conclude that the contribution of recombination to diversity is significant and positive, but recombination contributes little to divergence. This indicates that recombination is likely to modulate the footprint of selection in the genome. Next, we tested the impact of recombination rate on the efficiency of selection. We examined whether recombination rate (1) affects the distribution of nonsynonymous substitutions across the genome and (2) affects the pattern of diversity around nonsynonymous and synonymous substitutions. In particular, we use a generalized linear model to test how recombination modulates the magnitude and physical extent of the loss of diversity surrounding substitutions. Our analysis of these putative selective sweeps should be less sensitive to common confounding factors such as gene expression and GC content than previous measures. In total, this work allowed us to determine that recombination rate has an important impact on how selection shapes diversity across the genome of *Drosophila pseudoobscura* and its close relatives.

Materials and Methods

Fine-Crossover Maps: Crosses and Technical Details

Using a backcross design, we developed two recombination maps for *D. pseudoobscura*(Flagstaff and Pikes Peak) and one recombination map for *D. miranda*. For each cross, Duke's Genomic Analysis Facility genotyped 1,440 individual backcrossed flies for 384 line-specific SNP markers using the Illumina BeadArray platform (Illumina, San Diego, CA) (Fan et al. 2003).

Fine-Crossover Maps: Recombination Map Construction

Recombination events were scored when an individual fly's genotype changed from heterozygous to homozygous (for the parent in the backcross) or vice versa for autosomes and when the fly's genotype changed between the possible allele combinations for the sex chromosome arms XL and XR. Double crossovers were defined as adjacent intervals with different genotypes on both sides (for instance, a single homozygote genotype call nested in a tract of heterozygote genotype calls). We deemed these as genotyping errors as crossover interference is high within 2 Mb (Fitzpatrick 2009) and removed the single inconsistent genotype, scoring it as missing data. CentiMorgans were defined as the number of recombination events over the total number of individuals examined for each recombination interval, and we scaled this raw measure with a correction for recombination interference (Kosambi 1943). Throughout the article, recombination rates are given in Kosambi centiMorgans (Kosambi 1943)per Megabase (cM/Mb).

Approximately 1,400 backcross progeny were scored for the Pikes Peak *D. pseudoobscura*map, approximately 1,250 backcross progeny were scored for the Flagstaff *D. pseudoobscura*map, and approximately 1,170 backcross progeny were scored for the *D. miranda* map. Physical genomic distances used to calculate centiMorgans per Megabase (cM/Mb) per interval were based on the *D. pseudoobscura* reference genome v2.6 (Flagstaff) and v2.9 (Pikes Peak, *D. miranda*). Marker order was confirmed by the R (The R Foundation for Statistical Computing 2010) package OneMap (Margarido et al. 2007) using the algorithms Recombination Counting and Ordering (Van Os et al. 2005) and Unidirectional Growth (Tan et al. 2006). Onemap does not accommodate backcrossed designs for sex chromosomes; therefore, we specified an F2 intercross design in these cases. We found one small inversion in *D. miranda* relative to *D. pseudoobscura* on chromosome 2. We estimated the left breakpoint was between the markers at 10,491,527 and 10,660,216 bp, and the right breakpoint was between the markers at 13,318,705 bp and 14,068,383 bp from the telomeric end of chromosome 2. This inversion corresponds to one previously documented between *D. miranda* and *D. pseudoobscura* between markers *rosy* and *nop56* (Bartolome et al. 2006). Recombination rate differences are probably not due to differences in gene order; thus, we used the *D. pseudoobscura* orientation for this inversion when comparing recombination between maps and excluded intervals that included the breakpoints. Confidence intervals (95%) for cM/Mb for each recombination interval were calculated by permutation (Cirulli et al.

2007; Stevison et al. 2010). Confidence intervals for those intervals where we did not find a single recombinant individual were estimated from a binomial distribution—simply, we solved the equation $(1-x)^N = 0.05$, where x is the 95% upper bound of recombination frequency, and N is the number of individuals surveyed.

Fine-Scale Recombination Maps: Defining Intervals With Conserved and Divergent Recombination

The rationale for regressing out the effect of species (when identifying conserved intervals) was to account for the globally higher recombination rate in *D. miranda* relative to *D. pseudoobscura* and to identify regions where the recombination profile overlapped (e.g., where peaks and troughs can be overlaid). To delimit conserved regions using data that have not been corrected for elevated recombination rate of *D. miranda*, one might identify a region with very similar recombination rates between *D. miranda* and *D. pseudoobscura*, but this region may be a trough in recombination rate for *D. miranda* and a peak in recombination rate for *D. pseudoobscura*. Not correcting for the global elevation of *D. miranda* may lead to falsely concluding that a region has a conserved recombination profile between two maps. Thus, we used a rare events logistic regression (Zelig package in R) between each set of condensed fine-scale recombination maps to identify regions of conserved recombination after accounting for map identity (Flagstaff–Pikes Peak, Flagstaff–*D. miranda*, Pikes Peak–*D. miranda*). The package Zelig

uses the same model as a logistic regression, but it corrects for a bias that is introduced when the sample contains many more of one of the dichotomous outcomes than the other. Recombination events conditioned on the total number of observations was the response variable, and species, interval, and species-by-interval were included as factors in the model. We defined "divergent" intervals as those where tests in each interval between the species from the rare events logistic regression had a q-value of <0.05 after correction for multiple tests (Benjamini 1995). "Conserved intervals" were those intervals that displayed a nonsignificant difference across all three maps when analyzed with a rare events logistic regression and had an odds ratio between 0.62 and 1.615, after accounting for a species effect. We did not correct for multiple tests in defining conserved intervals.

In this way, only intervals that were conserved within and between species were delineated as conserved intervals. The final dataset used to differentiate between the mutagenic and selection hypotheses contained 27 conserved intervals on chromosome 2. We did not use the XR to differentiate between the mutagenic and selection hypotheses—of the 44 intervals condensed across three XR maps, only seven were conserved within and between species. We chose not to combine data from chromosome 2 and XR, as there is some evidence for different evolutionary patterns between autosomal and sex chromosomes in *Drosophila* (Vicoso et al. 2006).

Fine-Scale Recombination Maps: Recombination, Diversity and Divergence

We analyzed the effect of recombination on diversity and divergence by applying a quasibinomial GLM as the data were overdispersed, which has several statistical properties favorable to analyzing proportions such as pairwise diversity (Wilson 2002; Warton et al. 2011). Diversity or divergence was used as a response variable by binding the number of SNP bases to the number of non-SNP, eligible bases with cbind in R. We included recombination rate, proportion of G or C bases within the recombination interval, gene density (measured as a proportion of nucleotides within the recombination interval that are coding), a proxy for neutral mutation rate, and interaction terms as factors in the model.

For these models, the analysis presented is restricted to those conserved, condensed intervals with highly similar recombination rates between all three maps, unless otherwise noted. This restriction removes a classic bias by requiring that the intervals have similar recombination rates between the two species compared for the divergence measures (**Figure 1**). Similar linear models were also analyzed using the uncondensed intervals for each of the three maps individually. All statistics were performed in R version 2.12.1 (The R Foundation for Statistical Computing 2010) unless otherwise noted.

Ultrafine Crossover Maps: Recombination Map Construction and Analysis

Using Flagstaff 16 and Flagstaff 14, we followed the same backcross scheme described in the section "Fine-Crossover Maps: Crosses and Technical Details." Over 10,000 progeny from this backcross were stored in 96-well plates, frozen at −20°C and amplified for markers over these three regions. PCR products were visualized on a polyacrylamide gel using LICOR 4300.

Recombination and Nonsynonymous Substitutions

The number of nonsynonymous substitutions, specific to the *D. pseudoobscura*+*D. persimilis*lineage, were calculated for each gene using PAML using the resequenced genomic and reference genomic data (one *D. lowei*, three *D. miranda*, three *D. persimilis*, two *D. pseudoobscura bogotana*, and 11 *D. pseudoobscura* genomes, filtered for quality as described above). We used a tree rooted with *D. lowei* and considered the branches leading to [*D. persimilis* (*D. pseudoobscura*, *D. pseudoobscura bogotana*)] to be the foreground branches. We included *D. persimilis* a part of the foreground branch because relatively extensive interbreeding occurs between *D. pseudoobscura* and *D. persimilis* across much of the genome, aside from a few inverted regions (Machado et al. 2007; Noor et al. 2007; McGaugh et al. 2012).

Following (Bullaughey et al. 2008), we used a GLMM with Poisson distribution to examine the potential for recombination rate to shape the distribution of nonsynonymous substitutions along the *D. pseudoobscura*+*D. persimilis* lineage. The

model contained the following main effects: the number of silent segregating sites in each gene, GC content in each gene within Flagstaff 16, the proportion of coding bases 50 kb on either side of the gene's midpoint, weakly selected average pairwise divergence within the gene between *D. persimilis* and *D. lowei* at 4-fold degenerate sites of unpreferred codons (a proxy for neutral mutation rate), recombination rate observed for the interval containing the gene, and a random variable included to account for pseudoreplication of multiple genes per interval. The response variable was the number of nonsynonymous substitutions observed in each gene. This model construction allowed the inclusion of genes whose synonymous substitution count was zero (*sensu* (Bullaughey et al. 2008)). The GC content from Flagstaff16 was used as this was the line used for backcrossing in the crossing scheme, and the Flagstaff map (*D. pseudoobscura*) was used in this analysis.

Recombination and Reduction in Diversity Around Nonsynonymous Substitutions

We used 4-fold degenerate sites of unpreferred codons to measure the average levels of diversity as a function of distance from amino acid substitutions along the *D. pseudoobscura*+*D. persimilis* lineage (as identified by PAML, see above). Generalized linear mixed models with a Poisson distribution were used to compare the diversity around nonsynonymous substitutions along the *D. pseudoobscura*+*D. persimilis* lineage in relation to distance from the site and recombination rates measured

in the Flagstaff cross. Measures of diversity at 4-fold degenerate sites were taken 60 kb (*sensu* (Sattath et al. 2011)) from the site in either direction (120 kb total) with nonoverlapping bins of 1,000 bp. The random effects of identities of each substitution were estimated. We included as covariates (1) divergence between *D. persimilis* and *D. lowei* at 4-fold degenerate sites (a proxy for neutral mutation rate), (2) proportion of bases that were either G or C in Flagstaff 16 within the 1,000 bp window, (3) proportion of codons that were nonsynonymous substitutions within the 1,000 bp window, and (4) proportion of bases that were coding over each 1,000 bp window. The absolute value of the distance from the site and local recombination rate (at the particular nonsynonymous substitution) were included in the model as well as the interaction between distance and recombination rate. All effects in the model were standardized to mean zero and unit standard deviation. As a control, similar analyses were performed using synonymous substitutions along the *D. pseudoobscura+D. persimilis* lineage. Synonymous substitutions should evolve in a more neutral fashion; thus, less of an interaction between distance and recombination rate is expected. Any 1,000 bp window with less than 75 eligible, 4-fold degenerate sites was excluded from the analysis. Any nonsynonymous or nonsynonymous changes with less than 10 windows were excluded from the analysis. For the 60 kb analysis, after all filtering steps, our data consisted of 4,338 nonsynonymous and 8,670 synonymous substitutions along the *D. pseudoobscura+D. persimilis* lineage on chromosome 2. Four-fold degenerate sites were

used here, rather than 4-fold degenerate sites at unpreferred codons, because too little data were available in each 1,000 bp nonoverlapping window.

Results

We first discuss general features of the recombination landscapes we observed in *Drosophila pseudoobscura* and *D. miranda* before we address the implications of these observations for understanding diversity, divergence, and the nature of selection in the genomes we sequenced.

General Summary of Recombination Data: Fine-Scale Maps

We generated linkage maps for chromosome 2 and parts of the X chromosome for *D. pseudoobscura* and *D. miranda*. Using a backcross design and inbred lines, we developed two replicate recombination maps (referred to here as "Flagstaff" and "Pikes Peak") for *D. pseudoobscura* and one recombination map for *D. miranda* using the Illumina BeadArray platform to distinguish heterozygotes from homozygotes of the inbred lines used in the backcross design. These maps measure recombination rate across <200 kb windows, and we refer to these as "fine-scale" maps.

Recombination was surveyed across approximately 43% of the *D. pseudoobscura* physical genome and about 31% of the *D. miranda* physical genome. For each of the three maps, nearly the entire assembled region of chromosome 2 (97.8%–99.4%), the majority of the XR chromosome arm (70.8%–89.4%), and part of the XL

chromosome arm (~22%–23%) were surveyed. After removal of likely erroneous putative double recombinants, ambiguous genotypes, and markers that did not work or gave inconsistent genotypes, recombination was measured for three different crosses for 1,158–1,404 individuals per map. Excluding larger intervals at the telomeres and centromeres, intervals between markers had a median size across the three maps of 141–148 kb for chromosome 2 and 146–160 kb for the XR chromosome arm.

For chromosome 2, recombination rates ranged from 0–30.8 cM/Mb in *D. pseudoobscura* and 0–24.0 cM/Mb in *D. miranda*. The number of individuals surveyed is often slightly different per interval; therefore, for all intervals where no recombination was detected, we report 0 cM/Mb. The recombination rate for those intervals with "0 cM" should be interpreted as <1 recombination event per total number of individuals surveyed for each interval. Recombination near the telomere and centromere was measured at a broader scale than the remainder of chromosome 2 because we expected these regions to have lower crossover rates than the center of the chromosome (chromosome 2 is telocentric). Because of this limitation, comparisons of recombination rates between the ends of the chromosome and the center are more tentative. Nonetheless, examining recombination across roughly 3 Mb of sequence at the telomeric end and 3 Mb at the centromeric end, we found up to an 8.9-fold difference between the recombination rates at the middle of chromosome 2 relative to the centromeric end. The Pikes Peak *D. pseudoobscura* map exhibited the largest reduction of recombination at the

telomeric or centromeric ends relative to the center of the chromosome for all three maps, though in the Flagstaff *D. pseudoobscura* map and the *D. miranda* map, recombination rates were reduced by at least 2.6-fold in the centromere and telomere relative to the center of the chromosome.

For the XR chromosome arm, recombination rates ranged from 0–25.2 cM/Mb in *D. pseudoobscura* and 0–32.3 cM/Mb in *D. miranda*. The number of crossovers per individual for both chromosome 2 and the XR arm was close to 1 (1.01–1.06) for *D. pseudoobscura* and was 1.40–1.54 for *D. miranda*, illustrating that a greater overall recombination rate in *D. miranda*relative to *D. pseudoobscura* is observed in both an autosome and a sex chromosome.

The XL chromosome arm was not surveyed as intensively (~22%–23% of the XL arm in Pikes Peak and *D. miranda* and ~60% of the XL arm in Flagstaff). The number of crossovers per individual appears consistent with ~1 crossover per chromosome arm, as in *D. pseudoobscura* XR and chromosome 2, but the average number of crossovers per individual on the XL reflects how much of the arm was surveyed. For example, when ~22%–23% of the arm was surveyed, crossovers per individual ranged from 0.23–0.26. A binomial Generalized Linear Model (GLM) with size of the interval as a covariate and interval identity as a factor in the model indicated significant heterogeneity in recombination rate among intervals for chromosome 2, XR, and XL (each tested separately) for each of the three maps (each tested separately, interval

identity $p<0.00001$, $\chi^2 \geq 64.67$, $df \geq 3$, in all cases). Furthermore, 95% confidence intervals (generated via the same method in (Cirulli et al. 2007)) do not overlap in many cases between different intervals (shown in **Figure 3**). Overall, we observe heterogeneity in fine-scale recombination rates within each of the three maps (see **Figure 3** with 95% confidence intervals plotted), and we note a reduction in recombination rate around the telomeric and centromeric ends consistent with other studies in *Drosophila* (Stevison et al. 2010).

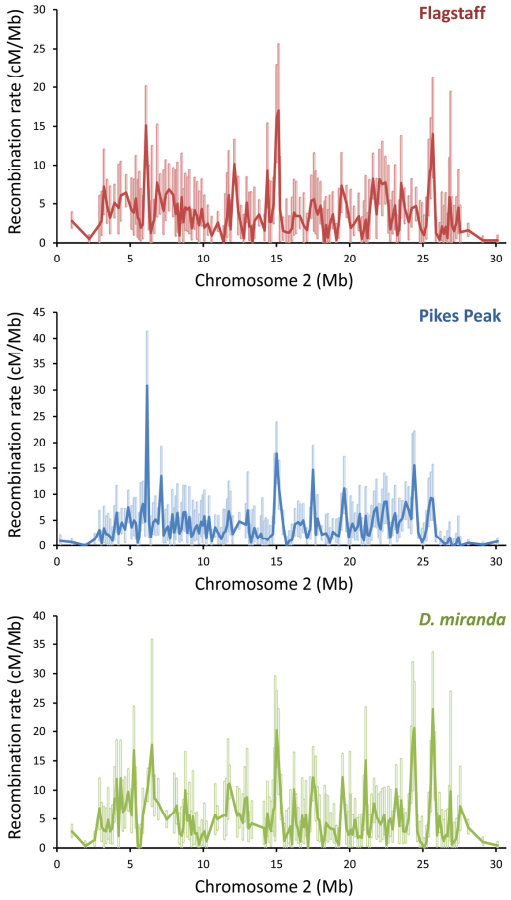

Figure 3: Fine-scale recombination rates on chromosome 2

General Summary of Recombination Data: Ultrafine-Scale Maps

Our three fine-scale crossover maps utilized markers on average 141–160 kb apart (median interval size for each of the three maps, with the exception of XL where the median distance between markers was 200–1,775 kb for the three crosses). We additionally examined three regions on chromosome 2 in more detail. Each of these regions spanned a total of 99–125 kb, and we placed markers every ~20 kb within the region (16 total intervals). These regions were originally picked because previous data (Kulathinal et al. 2008; Stevison et al. 2010) indicated that recombination rates for each of these regions differed (regions are referred to as 6 Mb, 17 Mb, and 21 Mb, which indicate approximate location on chromosome 2). We refer to these as "ultrafine-scale" maps. For these ultrafine maps, we followed the same backcross scheme as above, and we scored approximately 10,000 individuals for each marker. For the 16 ultrafine intervals, each interval was on average 20.61 kb long (range 12.6–27.4 kb).

Recombination rates range from 1.6–21.2 cM/Mb for these ~20 kb intervals (**Figure 4**). The ultrafine-scale map uncovered variation in recombination rates that was not apparent with the fine-scale maps. For example, for the 17 Mb ultrafine-scale region on chromosome 2, the recombination rates for the two fine-scale intervals spanning this region (17.5–17.7 Mb) are 5.6 and 4.4 cM/Mb. The ultrafine-scale recombination rates, in contrast, ranged from 3.5–21.2 cM/Mb (markers spanning 17.5–17.7 Mb). This heterogeneity in recombination rates within the ultrafine regions was statistically

significant (binomial GLM similar to that described in fine-scale section above: p = 0.0011, df = 14, χ^2 = 35.91) and highlights the fact that "broader" scale measures of recombination rates (such as the fine-scale measures here) are averages of true variation in recombination rate.

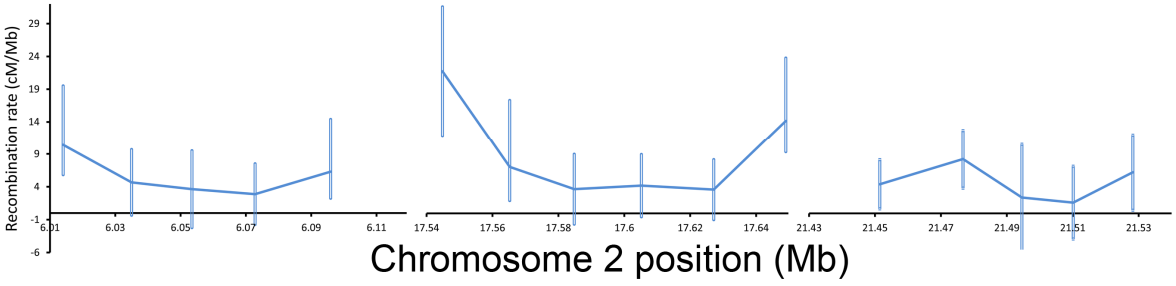

Figure 4: Ultrafine recombination rates

Recombination Rate Comparison between Maps

For comparisons of recombination rates between fine-scale maps, we restricted our analysis to intervals that were condensed to have nearly identical physical marker placement between the three fine-scale maps (**Figure 5**). Recombination was estimated as detailed above, using the number of crossovers spanning the newly defined physical intervals. After condensing across all three maps, 97 intervals remained for chromosome 2 and 44 intervals for XR (see **Table 4**). The XL chromosome arm was not included in the analysis that used condensed intervals across maps because too few intervals overlapped between all three maps. When comparing two maps, intervals were condensed between those two maps only.

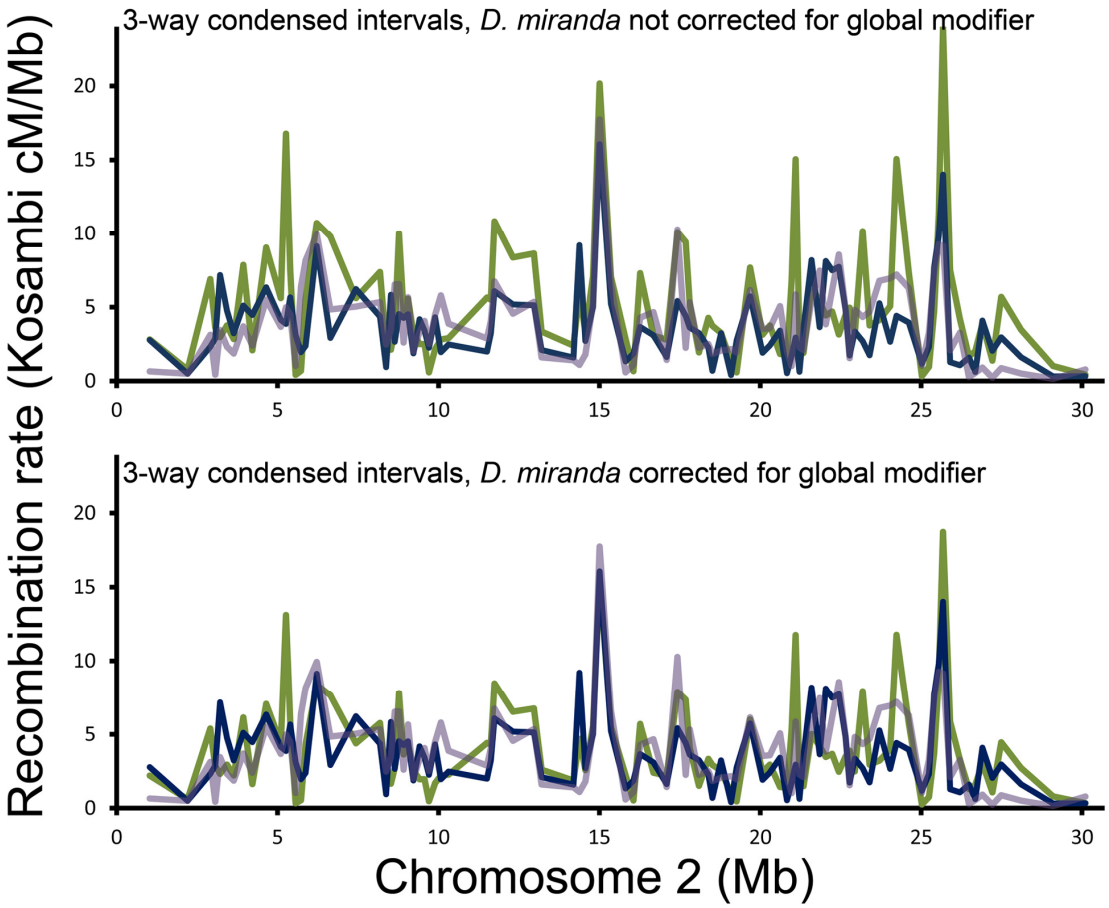

Figure 5: Fine-scale recombination rates for condensed intervals without and with global modifier correction

Local Recombination Rates between Two* D. pseudoobscura *Maps Are Similar

Recombination rates did not differ significantly between the two *D. pseudoobscura* maps for either the XR or chromosome 2 for the two-map comparisons (each chromosome analyzed separately, rare events logistic regression, absolute value of $z>0.3901$, $p>0.236$, in both cases). For chromosome 2, one interval was significantly different in recombination rate after correcting for multiple tests (Benjamini 1995). For the XR, no intervals between the two *D. pseudoobscura* maps were significantly different

in recombination rate after correcting for multiple tests. The 95% confidence intervals for the odds ratio of the difference between maps were narrow and located around zero, indicating that the maps are likely very similar (chromosome 2, 0.87–1.10; XR, 0.94, 1.28; within-species two map comparison). It is unlikely that the single significant difference observed within the same species is because of slight differences in marker placement between the two maps. The marker placement for this interval was nearly identical between the two maps (left marker, 102 nucleotides different between maps; right marker, 17 nucleotides).

Globally Higher Recombination Rate in D. miranda *Relative to* D. pseudoobscura

For both chromosome 2 and the XR chromosome arm, *Drosophila miranda* had significantly higher recombination rates than both *D. pseudoobscura* maps (**Figure 5**, **Table 4**). A rare events logistic regression of two-map comparisons indicated that the recombination rate of the *D. pseudoobscura* crosses we surveyed is about 76%–78% of the *D. miranda* recombination rate we observed on chromosome 2 (absolute z value>4.5374, p<0.001 for *D. miranda* relative to either *D. pseudoobscura* map, **Table 4**). The recombination rate of *D. pseudoobscura* is about 68%–71% of the *D. miranda* recombination rate on the XR chromosome arm (rare events logistic regression absolute z value>5.101, p<0.001 for *D. miranda* relative to either *D. pseudoobscura* map, **Table 4**).

Table 4: Comparison of intervals condensed within and between recombination maps

Map Comparisons		**Ch2 (N = 97)**	**XR (N = 44)**
D. pseudoobscura Pikes Peak-Flagstaff	different /conserved	1/60	0/21
	Odds ratio	0.9789 (0.8682, 1.1037)	1.0602 (0.8700, 1.2919)
	p-value	$p < 0.727$	$p < 0.562$
D. pseudoobscura Flagstaff - *D. miranda*	different /conserved	0/50	2/20
	Odds ratio	0.7794 (0.6916, 0.8787)	0.5860 (0.4877, 0.7041)
	p-value	$p < 0.001$*	$p < 0.001$*
D. pseudoobscura Pikes Peak - *D. miranda*	different /conserved	3/48	5/19
	Odds ratio	0.7629 (0.6780, 0.8584)	0.6213 (0.5267, 0.7328)
	p-value	$p < 0.001$*	$p < 0.001$*

Limited Local Recombination Rate Divergence between D. pseudoobscura *and* D. miranda

After the global difference between *D. miranda* and *D. pseudoobscura* is accounted for by the rare events logistic regression, recombination rates within and between species appear very similar for chromosome 2 (**Figure 5**). None of the intervals for the two-map comparison between *D. miranda* and *D. pseudoobscura*–Flagstaff were significantly different after correction for multiple tests, though power to detect significant differences on a per interval basis was likely weak. For example, 15 of the 115 intervals on chromosome 2 showed at least a 3-fold difference in recombination rate between maps, though this magnitude of difference was not significant in our rare events logistic regression after correcting for multiple tests. Likewise, only one of the

intervals for the two-map comparison between *D. miranda* and *D. pseudoobscura*–Pikes Peak was significantly different after correction for multiple tests, but 19 of the 123 intervals exhibited at least a 3-fold difference in recombination rate between maps for chromosome 2.

The XR chromosome exhibited a qualitatively larger difference in recombination rate between species than chromosome 2. After the global difference between *D. miranda* and *D. pseudoobscura* is accounted for by a rare events logistic regression, two of the intervals between *D. miranda* and *D. pseudoobscura*–Flagstaff for the two-map comparison and seven of the intervals between the *D. miranda* and *D. pseudoobscura*–Pikes Peak two-map comparison were significantly different after correction for multiple tests. Six of the 72 intervals between *D. miranda* and *D. pseudoobscura*–Flagstaff two-map comparison exhibited at least a 3-fold difference, and 12 of 102 intervals between *D. miranda* and *D. pseudoobscura*–Pikes Peak exhibited at least a 3-fold difference.

Twenty-seven of 97 condensed intervals (three-map comparison, condensed between all three maps) for chromosome 2 were considered to be "conserved" within and between species. This means that they displayed a nonsignificant difference across all three maps when analyzed with a rare events logistic regression and had an odds ratio between 0.62 and 1.615 after the effect of map identity was taken into account. These "conserved" intervals were used for further downstream analyses (see "Diversity, Divergence, and Recombination"). For the XR, seven of 44 intervals condensed between

all three maps were conserved within and between species according to the criteria outlined above.

In sum, we observe strong conservation in recombination rates within a single species, while between species, we see globally elevated recombination rates in *D. miranda*. Once the global difference is accounted for, there are few intervals with significant differences in recombination rate within and between species. Thus, it is possible and parsimonious that recombination rate is generally conserved at the scale examined here (~180 kb) over moderate evolutionary timescales (2–2.5 my).

Diversity, Divergence, and Recombination

We used various Illumina platforms to resequence genomic DNA from 10 *D. pseudoobscura*lines using virgin females from lines that were inbred for five or more generations with full-sibling single-pair mating. *Drosophila pseudoobscura* populations across North America display very little differentiation, as indicated by low F_{ST} values (always<0.10, often<0.05 for loci located outside of the inversion polymorphisms of the third chromosome) (Schaeffer et al. 1992; Noor et al. 2000). Therefore, the choice of strains sequenced for estimating diversity covered much of the species range but was fairly random. We also sequenced two lines of *D. persimilis* (one of these was provided by S. Nuzhdin), two lines of *D. pseudoobscura bogotana* (one of these was provided by S. Nuzhdin), one line of *D. lowei*, and three lines of *D. miranda* (two provided by D. Bachtrog, Short Read Archive accession numbers SRA044960.1, SRA044955.2, and

SRA044956.1; see also http://pseudobase.biology.duke.edu/). The divergence between *D. persimilis* and *D. lowei* was used to generate measures of a proxy for neutral mutation rate across the genome. In all diversity and divergence calculations, the reference sequences for the*D. pseudoobscura* and *D. persimilis* genomes were both included (Richards et al. 2005; Clark et al. 2007). Briefly, average pairwise diversity and divergence was calculated for 4-fold degenerate sites, focusing exclusively on unpreferred codons (Vicario et al. 2007), though we obtained very similar results when using all 4-fold degenerate sites. Overall, recombination is significantly and positively associated with average pairwise diversity but not average pairwise divergence at 4-fold degenerate sites of unpreferred codons. We examined this relationship in several ways.

Diversity, Not Divergence, Is Positively Associated with Recombination in All Intervals

We analyzed each chromosome for each uncondensed recombination map independently using a generalized linear model for diversity and a separate model for divergence. After accounting for multiple covariates, diversity at 4-fold degenerate sites of unpreferred codons shows a significant, positive relationship with recombination, while divergence at 4-fold degenerate sites of unpreferred codons does not. This result is consistent for each of the three recombination maps (*D. pseudoobscura*–Flagstaff, *D. pseudoobscura*–Pikes Peak, and *D. miranda*) for both chromosome 2 and the XR

chromosome arm. The XL chromosome arm contained too few intervals for analysis for*D. pseudoobscura*–Flagstaff. For *D. pseudoobscura*–Pikes Peak and *D. miranda*, diversity showed a significant, or nearly significant, positive relationship with recombination, while divergence did not.

Table 5: Factors affecting diversity within species at four-fold degenerate sites for unpreferred codons using intervals with conserved recombination rate

	Df	Deviance	Residual Df	Residual Dev.	F	*p*-value
Null			26	57.009		
Gene Density	1	2.3190	25	54.690	2.2948	0.147171
Mutation	1	12.7343	24	41.955	12.6013	0.002289*
Recombination	**1**	**6.1877**	**23**	**35.768**	**6.1231**	**0.023521***
GC	1	11.1854	22	24.582	11.0685	0.003751*
Gene Density*Mutation	1	2.0720	21	22.510	2.0504	0.169304
Gene Density*Recombination	1	2.8041	20	19.706	2.7748	0.113065
Mutation* GC	1	0.5488	19	19.157	0.5430	0.470669
Recombination* GC	1	0.0007	18	19.156	0.0007	0.978599

Table 6: Factors affecting divergence between species at four-fold degenerate sites for unpreferred codons using intervals with conserved recombination rate

	Df	Deviance	Residual Df	Residual Dev.	F	*p*-value
Null			26	53.578		
Gene Density	1	2.1647	25	51.414	1.1784	0.29201
Mutation	1	4.8404	24	46.573	2.6349	0.12192
Recombination	**1**	**0.2540**	**23**	**46.319**	**0.1382**	**0.71437**
GC	1	7.3218	22	38.997	3.9857	0.06124
Gene Density*Mutation	1	0.5094	21	38.488	0.2773	0.60492
Gene Density*Recombination	1	1.9069	20	36.581	1.0380	0.32178
Mutation* GC	1	0.0309	19	36.550	0.0168	0.89827
Recombination* GC	1	0.2399	18	36.310	0.1306	0.72202

Diversity, Not Divergence, Is Positively Associated with Recombination in Conserved Intervals

The analysis above suggests that the recombination–diversity relationship is probably the result of the effect of recombination on selection at linked sites; however, inadvertently including regions with discordant recombination rates between species in the analysis above could result in a pattern that supports this hypothesis—even when recombination is predominantly mutagenic (**Figure 1**). To resolve this potential bias, we restricted analysis to only regions that exhibited conserved recombination rates between all three chromosome 2 maps ($N = 27$ intervals; described above) and examined recombination in association with average pairwise *D. pseudoobscura* diversity at 4-fold degenerate sites of unpreferred codons (**Table 5**; **Figure 6**) and average pairwise *D. pseudoobscura–D. miranda* divergence at 4-fold degenerate sites of unpreferred codons (**Table 6**; **Figure 6**). The effect of recombination on diversity was significant when the analysis was restricted to only those regions with the most conserved recombination rates (quasibinomial GLM, $F = 6.123$, p value = 0.024), and the effect of recombination on divergence remained nonsignificant (quasibinomial GLM, $F = 0.138$, p value = 0.714). These regions contained only one interval within 4 Mb of the telomeric end and no intervals within 4 Mb of the centromeric end of the chromosome; thus, these results are not a function of broad-scale regional recombination rate differences across the chromosome. These results support the hypothesis that recombination affects diversity

through the effect of selection on linked sites. We did not perform an analysis on conserved windows for the X chromosome, as only seven intervals were conserved within and between species.

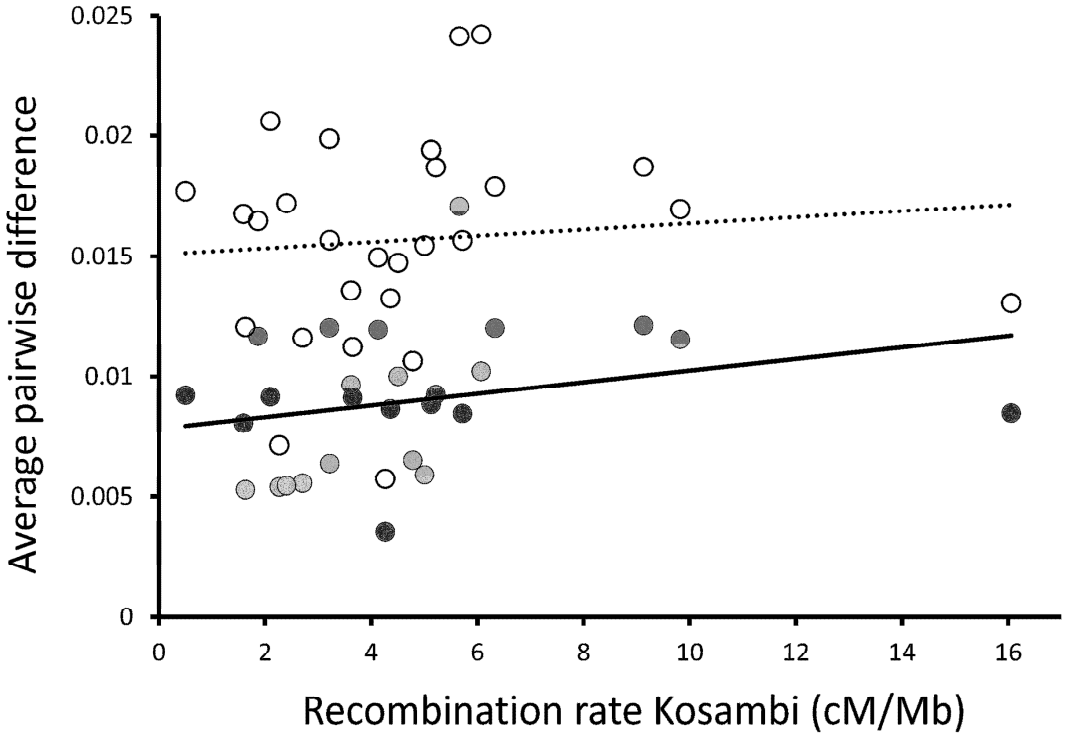

Figure 6: Relationship of recombination rate to diversity (filled circles, solid line) and divergence (open circles, dotted line) for fine-scale regions with conserved recombination between D. pseudoobscura–D. miranda.

Recombination and Selection

To determine the impact of recombination rate on selection at linked sites in the genome, we used two generalized linear models to analyze the relationship of recombination rate and several measures that may be indicative of the efficiency of selection: (1) abundance of nonsynonymous substitutions and (2) average pairwise

nucleotide diversity at 4-fold degenerate sites around nonsynonymous substitutions. We analyzed the association of recombination rate with these two measures in a generalized linear model framework to account for covariates such as gene density, GC content, and a proxy for neutral mutation rate. Biased gene conversion may influence substitution rates; thus, we controlled for GC content in all of the analyses below (Marais et al. 2001; Marais 2003; Duret et al. 2008; Webster et al. 2012). We did not consider gene expression as a covariate, though some studies point to a negative relationship with recombination rate (Haddrill et al. 2008).

No Correlation of Recombination With Nonsynonymous Substitution Abundance

The relationship of recombination rate to nonsynonymous substitution abundance was examined with the *D. pseudoobscura* Flagstaff fine-scale recombination maps. Nonsynonymous substitution abundance was measured as the nonsynonymous substitutions on the branch leading to *D. pseudoobscura*+*D. persimilis* as identified with PAML. The response variable was the number of nonsynonymous substitutions in each gene, and the covariates of the linear model included (1) the number of synonymous substitutions in the gene in question allowing for inclusion of genes where $K_s = 0$, (2), GC content of the gene, (3) gene density of 50 kb on either side of the midpoint of the gene, and (4) average pairwise divergence at 4-fold degenerate sites of unpreferred codons between *D. persimilis* and*D. lowei* as a proxy for neutral mutation rate within the gene.

We found no relationship (**Table 7**) between recombination and nonsynonymous substitution abundance with the fine-scale data (generalized linear model with Poisson distribution, $z = -0.614$, $p = 0.539$).

Table 7: Test for relationship between recombination rate and number of nonsynonymous substitutions

Response: Nonsynonymous substitutions along the *D. pseudoobscura*+ *D. persimilis* lineage					
Model	**Factor tested**	**Estimate**	**Std. error**	**z-value**	***p*-value**
Fine-scale	(Intercept)	2.574891	0.207963	12.38	<0.0001*
	Synonymous	0.053427	0.001557	34.31	<0.0001*
	GC content	-4.892668	0.339146	-14.43	<0.0001*
	Gene density	0.158809	0.196072	0.81	0.418
	Neutral mutation rate	0.470959	3.417997	0.14	0.890
	Recombination	**-0.015829**	**0.019014**	**0.83**	**0.405**

Footprints from Putative Hitchhiking May Be Slightly Larger in Low Recombination Regions

In response to selective sweeps, a trough in diversity should be visible around selected variants (Kaplan et al. 1989; Stephan 1992; Kim et al. 2002; Stephan 2010; Sattath et al. 2011). We analyzed diversity surrounding the nonsynonymous substitutions along the lineage leading to *D. pseudoobscura*+*D. persimilis* identified by PAML. We compared the average pairwise diversity patterns at 4-fold degenerate sites surrounding these substitutions in relation to the Flagstaff recombination rate and distance in basepairs from the substitution. In regions with high recombination rates, the footprints of selection are thought to be narrower than in regions with low recombination rates, where strong linkage between sites will create a stronger signature of sweeps (Kaplan et

al. 1989; Kim et al. 2002; JH 2004; Lohmueller et al. 2011). As a control, similar analyses were performed using synonymous substitutions along the *D. pseudoobscura*+*D. persimilis* lineage following (Sattath et al. 2011). Synonymous substitutions, in many cases, evolve in a more neutral fashion than nonsynonymous substitutions ((Sattath et al. 2011), but see (Kern et al. 2002; Bartolome et al. 2005)). In a recent genome-scale analysis conducted with data similar to what are presented here, little reduction in diversity was seen around synonymous substitutions (Sattath et al. 2011); this study instead saw an increase in diversity, which disappeared after correction for local mutation rates.

We considered 60 kb on either side of the substitution along the *D. pseudoobscura* lineage divided into 1,000 bp nonoverlapping windows (*sensu* (Sattath et al. 2011)). For each 1,000 bp window, the response variable was the number of polymorphic 4-fold degenerate sites. The generalized linear model included the following covariates: (1) total 4-fold degenerate sites, (2) GC content, (3) proportion of coding bases, (4) divergence of *D. lowei–D. persimilis* at 4-fold degenerate sites as a proxy for neutral mutation rate, and (5) proportion of bases that were nonsynonymous substitutions. The identities of each nonsynonymous substitution were included as random effects. This generalized linear mixed model with Poisson distribution included the following factors: absolute physical distance from the substitution, fine-scale-derived estimates of recombination rate, and the interaction between these two factors. A

negative interaction term means that short distances from a substitution and high recombination rates have similar effects on diversity as large distances and low recombination rates. We expect the interaction term for distance and recombination rate to be much reduced in magnitude for synonymous substitutions in comparison to the nonsynonymous analysis.

We found a small but significant negative interaction term of physical distance from the nonsynonymous site and recombination rate on nucleotide diversity around nonsynonymous substitutions (Poisson GLMM, $z = -7.52$, $p<0.001$; **Figure 7** and **8**). In other words, higher rates of recombination allow for recovery of diversity at shorter physical distances from the nonsynonymous site than lower recombination rates (**Figure 8**). In contrast, a weaker interaction was detected for the interaction of distance and recombination rate on diversity around synonymous substitutions along the *D. pseudoobscura* lineage (Poisson GLMM, $z = -2.43$, $p = 0.015$; **Figure 7** and **8**). GLM plots for the very low recombination rates of <0.5 cM/Mb show wider dips in diversity (and more associated noise; **Figure 8**) than plots for recombination rates of >0.5 cM/Mb (**Figure 8**).

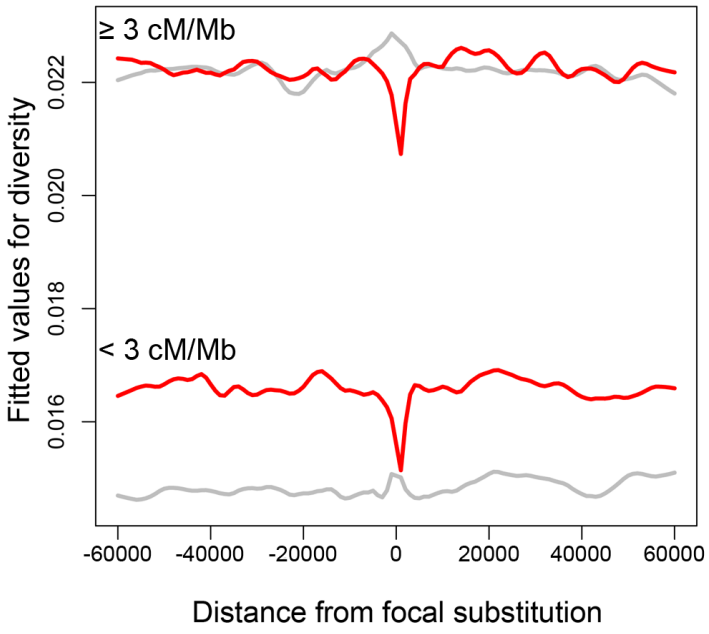

Figure 7: Footprints of diversity around substitutions

Distance from a substitution had a positive, significant effect on diversity as expected if linked selection of substitutions generates a dip in diversity. Recombination rate also had a positive, significant effect on diversity as expected, if either recombination was mutagenic or if positive/negative selection was operating on the chromosome. The proportion of nonsynonymous substitutions around a substitution had a negative significant effect on diversity surrounding a nonsynonymous site as expected if many of these substitutions combine forces to generate stronger selective sweeps. The interaction term pointing to deeper dips in diversity for lower recombination rates is no longer significant when examining only 5 kb or 15 kb on either side of the focal substitution (it is negative for nonsynonymous substitutions and

positive for synonymous substitutions), but it is conceivable that this lack of significance represents an issue with window size or sampling.

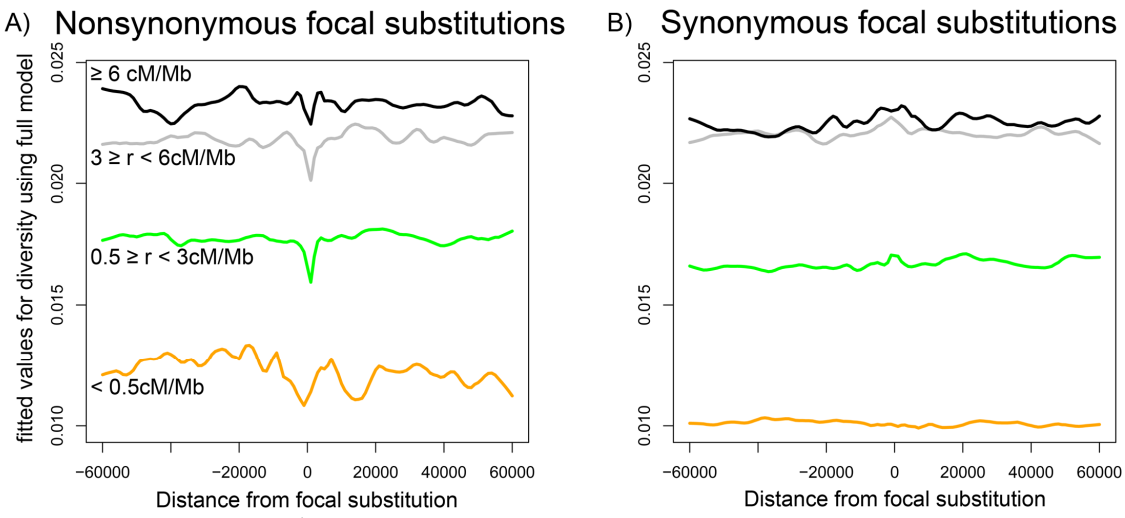

Figure 8: Footprints in diversity around substitutions

Discussion

Overall, our study identified both global and local differences in recombination rate between two closely related species of *Drosophila*. Aside from regions with exceptionally low recombination rates (Haddrill et al. 2007; Betancourt et al. 2009), variation in local recombination rates between species must be accounted for prior to concluding that the association between recombination rate and diversity is probably caused by recombination modulating the effects of selection at linked sites (Charlesworth 2010). By restricting our analysis in the *Drosophila pseudoobscura* system to only those regions with conserved recombination rates within

and between species, we rejected the hypothesis that recombination rate (at the scale tested) significantly affects divergence at 4-fold degenerate sites for unpreferred codons. These results support the conclusion that recombination has a substantial impact on how selection affects diversity in the genome. Furthermore, additional analyses suggest that recombination rate variation affects the impact of Hill–Robertson effects like selective sweeps and background selection in this system.

Ultrafine and Fine-Scale Variation in Crossover Rate in Drosophila

Here and in other recent work (Cirulli et al. 2007), we demonstrate that ultrafine-scale patterns of crossover rate (intervals spanning 20 kb) are also significantly heterogeneous in *D. pseudoobscura*. In each ultrafine region on chromosome 2, recombination rates varied by up to 6-fold (17 Mb region) over only approximately 120 kb (6 Mb region variation is 3.6-fold, and 21 Mb region variation is 5.1-fold), and ultrafine-scale maps reveal variation not detected in the fine-scale maps. This was especially apparent for the 17 Mb region, where ultrafine-scale recombination rates ranged from 3.5 to 21.2 cM/Mb, and fine-scale recombination rates in the same area ranged only from 4.4 to 5.6 cM/Mb. This heterogeneity suggests that our fine-scale measures (intervals spanning <200 kb) are averages of actual variation in recombination rate.

In humans, broad-scale variation averages over the density and intensity of ~2 kb hotspots that occur in clusters every 60–90 kb (McVean et al. 2004; Myers et al. 2005). The majority of recombination occurs at these hotspots, and the majority of recombination is governed by the DNA binding protein PRDM9 and its recognition motifs in humans (Baudat et al. 2010; Berg et al. 2010; Hochwagen et al. 2010; Parvanov et al. 2010; Hinch et al. 2011). Interestingly, several studies in different regions of the *D. melanogaster* genome indicate that linkage disequilibrium decays rapidly (Langley et al. 2000; Schaeffer et al. 2001; Palsson et al. 2004; Mackay et al. 2012), suggesting that the heterogeneity we observed in ultrafine-scale maps may not be governed by clustered hotspots similar to those in humans, or at least that a nontrivial amount of recombination may occur outside such "hotspots."

To assess whether "hotspots" of some sort exist in *D. pseudoobscura*, genome-wide patterns of linkage decay need to be investigated or incredibly fine-scale maps (interval size <5 kb) need to be made. Such a line of inquiry would help address basic questions about the requirements for functional recombination across various taxa. For example, there are several notable differences regarding the formation and function of the synaptonemal complex and the role of double-strand breaks across taxa (Hawley et al. 1992; Roeder 1997; Keeney 2001; Hawley et al. 2002; McKim et al. 2002; Page et al. 2003). Furthermore, the *Drosophila* lineage completely lacks several proteins essential for generating crossovers and double-strand break repair in other organisms (McKim et al.

2002; Schurko et al. 2008). It is likely that understanding particular sequence features associated with recombination on a kilobase scale in *Drosophila* will uncover more details about the mechanistic underpinnings of meiosis that differentiate these species and the distribution of crossovers across the genome.

Drosophila miranda *Has Elevated Global Recombination Rate Relative to* D. pseudoobscura

Recombination rates at broad scales are conserved between populations and species (Brooks et al. 1986; True et al. 1996; Backstrom et al. 2010; Beukeboom et al. 2010; Meznar et al. 2010; Dumont et al. 2011; Smukowski et al. 2011). Our fine-scale data are generally consistent with these findings except that *D. pseudoobscura* has about three-fourths the rate of recombination, on average, as *D. miranda* for chromosome 2 and about three-fifths the rate of recombination of *D. miranda* on the XR chromosome arm. Notably, *D. melanogaster* has one of the lowest recombination rates in the genus, as evidence indicates that *D. mauritiana*, *D. simulans*, *D. virilis*, *D. pseudoobscura*, *D. miranda*, and *D. persimilis* all exhibit higher rates of recombination (True et al. 1996; Ortiz-Barrientos et al. 2006; Stevison et al. 2010); this should be considered when interpreting hitchhiking and linkage data from *D. melanogaster* to patterns of recombination in *Drosophila* in general.

Recombination Prevents Diversity Erosion During Selection

Our results indicate that recombination affects diversity through mediating selection in the genome. While accounting for multiple covariates, we found no association between recombination and average pairwise divergence at 4-fold degenerate sites of unpreferred codons, and a significant, positive association of recombination with average pairwise diversity at 4-fold degenerate sites of unpreferred codons. Using data from our fine-scale maps, we ensured that recombination rates are nearly identical between the species used to generate divergence estimates; thus, we absolved a key assumption made in previous studies (see **Figure 1**). Data from *Drosophila* suggest both positive and negative selection are markedly less efficient in nearly nonrecombining regions of the genome (Bachtrog et al. 2002; Betancourt et al. 2002; Haddrill et al. 2007; Betancourt et al. 2009; Campos et al. 2012), and a relationship of diversity but not divergence to recombination is apparent for other species of *Drosophila* (Begun et al. 1992; Presgraves 2005; Begun et al. 2007; Stevison et al. 2010), mouse (Takahashi et al. 2004), beet (Kraft et al. 1998), tomato (Stephan et al. 1998; Roselius et al. 2005), *Caenorhabditis* (Cutter et al. 2010), and yeast (Cutter et al. 2011). This last example is especially interesting because recombination is known to be mutagenic in yeast (Strathern et al. 1995; Rattray et al. 2001), but there is a negative or absent divergence–recombination correlation (Noor 2008; Cutter et al. 2011); thus, it may be that recombination is somewhat mutagenic in many organisms, but the power of

recombination to modulate the diversity eroding effects of selection likely has a much greater impact on the genome.

In other systems, the divergence–recombination association is positive, which may be interpreted as evidence that recombination is predominately mutagenic. A positive divergence–recombination association is apparent for humans (Hellmann et al. 2003; Spencer et al. 2006), maize (Tenaillon et al. 2004), and in an inverted region between *D. pseudoobscura* and *D. persimilis* (Kulathinal et al. 2008). This association may be attributable to mutation (Lercher et al. 2002), but unmeasured variables or segregating ancestral polymorphism could predispose a system to exhibiting a positive divergence–recombination relationship (Begun et al. 2007; Noor 2008; Noor 2008; Cutter et al. 2010; Lohmueller et al. 2011). For instance, in *C. briggsae*, segregating ancestral polymorphism leads to the signature of recombination-associated mutation (i.e., a positive divergence–recombination association), but further examination shows the majority of polymorphism heterogeneity is caused by recombination affecting the impact of selection at linked sites (Cutter et al. 2010).

Recombination Rate and Abundance of Nonsynonymous Substitutions

Since recombination probably mediates the effects of hitchhiking in our system, we sought to understand whether this hitchhiking is primarily positive or negative (background, purifying) selection and if recombination rate variation has a significant

impact on the potential efficacy of selection. Evidence is emerging that in many organisms, especially those with large population sizes, selection may play a substantial role in shaping the genome (Hahn 2008). For partial selfers, it seems that background selection substantially affects the genome (Cutter et al. 2003; Nordborg et al. 2005; Rockman et al. 2010; Slotte et al. 2011), while in outcrossing species *Drosophila*, mice, and *Capsella grandiflora* a large fraction of the genome may be influenced by positive selection (Sella et al. 2009; Halligan et al. 2010; Slotte et al. 2011). The majority of studies find strong support that recombination can shape adaptive evolution when comparing regions of no recombination to regions with some or abundant recombination. However, after accounting for multiple covariates in regions with detectable recombination rates, there is often very little relationship between recombination rate and the efficacy of selection (Marais et al. 2001; Haddrill et al. 2007; Webster et al. 2012).

Across chromosome 2, we found no relationship between the number of nonsynonymous substitutions and the recombination rate as measured with our fine-scale Flagstaff map. Reanalysis of the fine-scale data after removal of the first and last 3 Mb of the chromosome did not change the relationship of fine-scale recombination rate to nonsynonymous substitutions.

Recombination Rate and Diversity Around Nonsynonymous Substitutions (GLM)

Our observation of a reduction of average pairwise diversity at 4-fold degenerate sites around nonsynonymous substitutions (**Figure 8**) is consistent with the idea that positive selection may have fixed many nonsynonymous substitutions along the ancestral lineage leading to *D. pseudoobscura+D. persimilis*, as has been argued elsewhere for other *Drosophila* species (Haddrill et al. 2010; Sattath et al. 2011). While potentially less common, dips in diversity could also be caused by deleterious mutations that can get fixed by chance if deleterious selection coefficients are small enough—a situation we call "loser's luck" (but see (Loewe et al. 2006; Haddrill et al. 2010)), and theoretical investigations of entirely neutral substitutions showed that their quick fixation can also lead to dips in diversity (Tajima 1990). Thus, while many of the dips in diversity we see may be caused by positive selection, both loser's luck and fixation of neutral substitutions may also contribute.

Diversity may be recovered slightly farther from a nonsynonymous substitution in areas of low recombination than in areas of high recombination, and such a relationship is not as pronounced for synonymous substitutions fixed along the lineage leading from the common ancestor of *D. pseudoobscura* and *D. persimilis* (**Figure 8**). Similarly, in *Arabidopsis*, haplotype blocks around nonsynonymous SNPs are larger than around synonymous SNPs (Kim et al. 2007). Our data agree with theoretical expectations (Kaplan et al. 1989; Kim et al. 2002) and past studies that show negative correlations of polymorphisms and nonsynonymous substitutions in

Drosophila (Andolfatto 2007; Begun et al. 2007; Macpherson et al. 2007; Sattath et al. 2011); indeed, our data also show a significant negative relationship for nonsynonymous substitutions and within-species polymorphisms, generally. Yet the negative interaction term between recombination rate and distance from focal substitutions we observed is dependent on window size and distance from the substitution examined.

Conclusions

Our study documented global and local differences in recombination rate between two closely related species, and these data indicate that recombination probably modulates Hill–Robertson effects in the genome, causing a positive association of diversity with recombination. While we found no overall association of recombination rate with the number of nonsynonymous substitutions at the fine scale, we found evidence for dips in diversity around nonsynonymous substitutions that are dependent on the distance from the substitution, local recombination rate, and a number of other factors. In total, our study adds to the growing literature that indicates that selection must be a ubiquitously important factor for shaping diversity across much of the genome (Smith et al. 1974; Kaplan et al. 1989; Kim et al. 2002).

3. Zinc finger binding motifs do not explain recombination variation within or between species of Drosophila

Caiti S.S. Heil and Mohamed Noor

Citation: Heil, C. S. and M. A. Noor (2012). "Zinc finger binding motifs do not explain recombination rate variation within or between species of Drosophila." Plos One **7**(9): e45055.

Introduction

Meiotic recombination is an essential process both mechanistically and evolutionarily, and thus should experience strong selective pressures. However, identifying how selection affects the locations of recombination events is more complex than was once assumed. Recombination rate is variable within and among genomes, displaying significant heterogeneity across most living organisms and evolving rapidly, with recombination "hotspot" turnover in as short as 120,000 years (Jeffreys et al. 2009). While years of research have determined some elements associated with recombination rate variation, such as temperature, GC content, repeats, SNP density, chromatin state, and histone modifications, the specific effects of DNA sequence "motifs" have attracted much attention.

The 13-base degenerate motif CCNCCNTNNCCNC recruits recombination events in approximately 40% of European human hotspots (Myers et al. 2008; Baudat et al. 2010). This motif binds the Cys_2His_2 zinc finger protein PRDM9 in humans, and allelic variation at *Prdm9* modifies hotspot activity within both humans and mice (Baudat et al.

2010; Berg et al. 2010; Parvanov et al. 2010; Grey et al. 2011; Hinch et al. 2011). The *Prdm9* gene also contains a SET-methyltransferase domain, which is responsible for the common chromatin feature trimethylation of lysine 4 of histone H3, or H3K4me3. H3K4me3 in yeast and mouse seems to be a prominent and pre-existing mark of active recombination sites, creating a link between sequence and epigenetic features affecting recombination (Borde et al. 2009). This link inspired a proposed model in humans involving the recognition of a DNA sequence motif by PRDM9 and the modification of adjacent nucleosomes by the SET domain (Barthes 2011). Proteins with an affinity to the modification H3K4me3 are recruited and may modify the chromatin or nucleosomes further. The conserved topoisomerase II- like protein SPO11 subsequently recognizes one or several of these signals, binds to the DNA at that location, and initiates recombination by a double-strand-break (DSB).

Cys_2His_2 zinc fingers are among the most common DNA-binding motifs found in eukaryotic transcription factors. These zinc finger proteins usually contain multiple "fingers", all of which have a conserved $\beta\beta\alpha$ structure with amino acids in the α-helix

contacting DNA in the major groove of the double helix (Wolfe et al. 2000) (**Figure 9**).

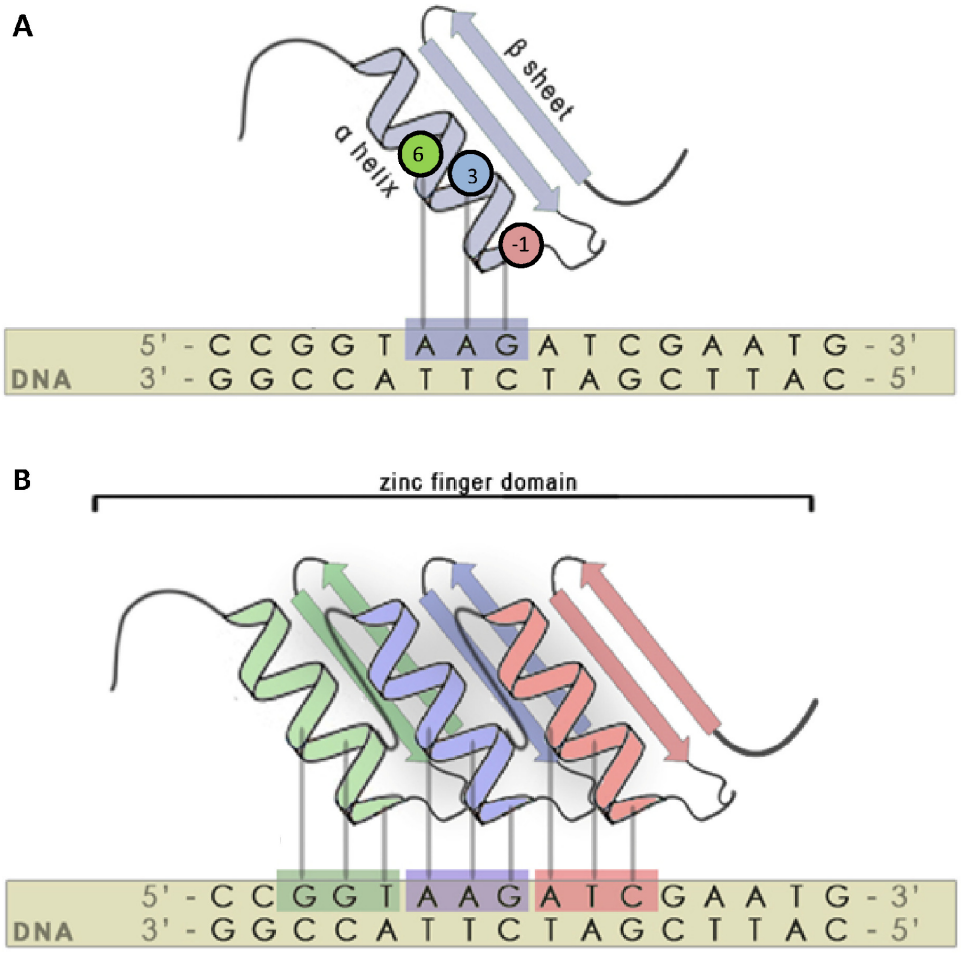

Figure 9: A model of Cys2His2 zinc finger binding

Zinc finger proteins function chiefly in protein-DNA binding, but also may be involved in protein-RNA binding and protein-protein binding, making them key elements in transcriptional regulation and many other processes. While transcription factors have long been recognized for their required role in yeast α recombination hotspots (Petes 2001), the discovery of *Prdm9* is the first implication of zinc finger

proteins and their predicted binding sequence motifs as major determinants of recombination hotspot location and usage in multi-cellular organisms (Segurel et al. 2011).

Yeast and mammals share many conserved meiotic proteins and processes. However, the Drosophila meiotic recombination process differs from yeast and mammals in several key components. First, homologous chromosome pairing and synapsis proceed normally in the absence of double strand breaks in Drosophila (Hawley et al. 1992; Roeder 1997; Keeney 2001; Hawley et al. 2002; McKim et al. 2002; Page et al. 2003). Indeed, Drosophila males undergo meiosis without any homologous recombination, a phenomenon rarely seen in other animals (Morgan 1910; Rasmussen 1977). Furthermore, the synaptonemal complex, a proteinaceous structure that binds homologs together during meiotic prophase, is conserved in structure but has diverged in function in Drosophila. The Drosophila synaptonemal complex does not require SPO11 to form, and functions in the initiation of recombination and the facilitation of DSBs. Finally, Drosophila lack some genes known to be crucial in recombination in other organisms, like *Msh4* and *Msh5*, and use a smaller subset of proteins in DNA repair (McKim et al. 2002; Schurko et al. 2008).With known differences in meiotic proteins and some apparent differences in the initiation of recombination, it is unclear if Drosophila would possess a recombination initiation process involving a protein like PRDM9.

In 2011, Lake *et al.* demonstrated that the Cys_2His_2 zinc finger protein *trade embargo* is required for meiotic DSB initiation in *Drosophila melanogaster* and suggested its essential role for processing DSBs into crossovers (Lake et al. 2011). However, while *Prdm9* binds to discrete sites across the genome, *trade embargo* appears to bind the entire length of the chromatin, casting doubt on the similarity between the two proteins. Nonetheless, the discovery of *trade embargo*'s role in DSB initiation and resolution implicates a general role for zinc finger proteins in the distribution of recombination.

Here, we explore the importance of Cys_2His_2 zinc finger genes in the initiation of Drosophila recombination and whether the abundance of predicted binding sites of such genes may correlate with recombination variation within and among species. First, we attempt to identify a *Prdm9* homolog in Drosophila, and confirm that *Prdm9* is indeed missing in this lineage. We then attempt to characterize any other zinc finger proteins involved in recombination by analyzing associations between predicted DNA sequence motifs and our empirically derived broad-and fine-scale measures of recombination rate in *D. pseudoobscura* and its close relative. As a validation of our approach, we apply the same procedure to the *Prdm9*-predicted motif to comparably scaled measures of recombination rate in humans. Our results suggest that Drosophila possess a recombination initiation mechanism disparate from human *Prdm9*.

Materials and Methods

System and Datasets

The species *Drosophila pseudoobscura* was selected due to the availability of high resolution recombination data not yet available in *Drosophila melanogaster*. Additionally, the availability of recombination data in closely related species *D. miranda* facilitates inter-specific comparisons. *D. pseudoobscura* inhabits the western coast of North America and diverged from *D. melanogaster* approximately 55 mya (Tamura et al. 2004). The recombination datasets for *D. pseudoobscura* consist of two recombination maps from the Flagstaff population (collected Flagstaff, AZ 1997), for more information about the recombination maps, see McGaugh *et al.* (2012)(McGaugh et al. 2012). The "superfine" recombination map consists of three ~100kb regions on Chromosome 2 with markers spaced within these regions every 20kb (219 crossovers observed). Briefly, the map was constructed by genotyping over 10,000 F2 backcross progeny by PCR at 19 markers. Their coordinates on chromosome 2 are: 6.003 Mb- 6.108 Mb (6 markers, 5 intervals, average interval 20.280 Kbp), 17.534 MB - 17.660 MB (7 markers, 6 intervals, average interval Kbp 20.878), 21.438 Mb -21.537 Mb (6 markers, 5 intervals, average interval 19.870 Kbp). The "broad-scale" recombination map spans the majority of Chromosome 2 with markers approximately every 180 kb, resulting in 140 intervals (1344 crossovers observed) (McGaugh et al. 2012). The broad-scale map was constructed by genotyping

1440 individual backcrossed flies for 384 line-specific SNP markers using the Illumina BeadArray platform (Fan et al. 2003) (Illumina, San Diego, California, United States). For comparisons across species, a "broad-scale" *D. miranda* recombination map of Chromosome 2 was used. The *D. miranda* broad-scale map was constructed using the same method as the *D. pseudoobscura* broad-scale map, and SNP markers were designed at comparable physical coordinates. For the between-species comparison, both the *D. pseudoobscura* and *D. miranda* recombination maps were condensed to comparable interval sizes, yielding 97 windows of about 320kb (McGaugh et al. 2012). Chromosome 2 is 30 Mb and makes up 23% of the physical genome.

DNA sequence for the strains corresponding to the recombination maps was also obtained from McGaugh *et al.* (2012)(McGaugh et al. 2012). We obtained the amino acid sequence for Cys_2His_2 zinc finger proteins for *D. melanogaster* and *D. pseudoobscura* from FlyBase (Lyne et al. 2007; McQuilton et al. 2012), for *D. persimilis* from FlyMine (Lyne et al. 2007), and for *D. miranda* from our own sequence data (McGaugh et al. 2012).

Identification of a Prdm9 homolog using BLAST

We used NCBI BLAST protein tools blastp and PSI-BLAST and the nucleotide tool blastn with default parameters, specifying the organism as Drosophila (Altschul et al. 1990). For the input query, we examined all genes and proteins annotated as *Prdm9*, selecting

Homo sapiens, *Strongylocentrotus purpatus*, and *Mus musculus* PRDM9 proteins and *Prdm9* sequence for input queries.

Motif Prediction

We used custom Perl and Unix scripts to extract the zinc finger domains from each *D. pseudoobscura* protein using the canonical Cys_2His_2 binding pattern CX(2-6)CX(11-13)HX(2-6)H. Each Cys_2His_2 protein contains a number of zinc fingers ranging from one to 21, with an average of five. We used a protein only if it had more than one zinc finger, obtaining binding sequences longer than 3 base pairs. This procedure resulted in an amino acid dataset of 186 unique *D. pseudoobscura* proteins. For each zinc finger, we recorded the amino acid residues at positions -1, 3, and 6 in relation to the start of the alpha helix, which are responsible for predicting DNA binding specificity (Kaplan et al. 2005). To generate the DNA sequence that these amino acids are predicted to bind to, we used two approaches.

To examine candidate proteins containing a SET domain, those identified in our BLAST searches, or ontologically identified as functioning in meiosis, we used the rigorous approach of Baudat *et al.* (2010) to generate the sequence motif for *Prdm9*. Briefly, we used the Zinc Finger Consortium database to obtain a matrix of binding residues, positions, and empirically determined binding sequence (Fu et al. 2009), then input this data into WebLogo to generate the sequence motif (Crooks et al. 2004). To take

into account that not all zinc fingers may be important in binding, we used a 3 letter sliding window for the DNA sequence motifs, looking at the whole motif and all possible contiguous 9 bp motifs from the whole motif. This approach was applied to zinc finger proteins GA18168 (*trade embargo*), GA23469 (*Blimp-*1), GA25755 (*hamlet*), GA26409 (CG9817), GA25849 (*crooked legs*), GA26228 (CG5245), GA26117, GA21024 (*combgap*), GA21437 (*teflon*), and GA17308 (*grauzone*) (**Table 8**).

For all other zinc finger proteins, we used the more scalable program enoLOGOS, with default parameters (Workman et al. 2005). The input for this program simply requires the amino acid contact residues for each zinc finger. The output is a normalized sequence logo of nucleotides, with the information content of each nucleotide position measured in bits (ranging from zero to two). Again, we used a 3 letter sliding window for the DNA sequence motifs, looking at the whole motif and all possible contiguous 9 bp motifs from the whole motif.

Motif occurrence

DNA sequence for *D. pseudoobscura* Flagstaff was split into intervals of known recombination. This resulted in 140 windows of average size 180kb for the *D. pseudoobscura* Flagstaff broad-scale dataset and 16 windows of average size 20kb for the *D. pseudoobscura* Flagstaff superfine-scale dataset. To identify the frequency of occurrence of all *D. pseudoobscura* zinc finger motifs, we used the EMBOSS command

"dreg"(Rice et al. 2000). The command "dreg" searches one or more sequences with the supplied regular expression and writes a report file with the matches. The frequency of motifs in a given interval for forward and reverse strands was combined and corrected for interval size, then regressed with recombination rate using custom Perl and R scripts. p-values were adjusted for multiple comparisons using a sequential Bonferroni correction (Rice 1989). For proteins that were significantly associated with recombination after correction for multiple comparisons, we ran a multiple regression accounting for GC content (JMP Version 9.0. SAS Institute Inc., Cary, NC).

Amino acid differences between species

To identify changes in Cys_2His_2 zinc fingers that alter DNA binding, we compared number of fingers and amino acids at positions -1, 3, and 6 for each protein in *D. melanogaster*, *D. miranda*, and *D. persimilis* to *D. pseudoobscura* using a custom Perl script. After identifying proteins that had differences in their zinc fingers between *D. pseudoobscura* and *D. miranda*, we followed the protocol outlined in the Motif Occurrence section above, but using the condensed *D. miranda* recombination data and sequence with this subset of proteins. The frequency of predicted motifs for this subset of *D. miranda* zinc finger proteins was identified using *D. miranda* recombination and sequence and *D. pseudoobscura* recombination and sequence, and then the correlation

coefficients compared. The same was done for predicted motifs for the same subset of *D. pseudoobscura* zinc finger proteins.

Candidate motif analysis

To identify any overrepresented sequence motifs not *a priori* associated with zinc finger binding, we used the EMBOSS command "wordcount," which counts and extracts all possible unique sequence words of a specified size in one or more DNA sequences. This analysis was done using a word size of six with the superfine-scale and broad-scale recombination datasets. To identify associations with recombination rate, the forward and reverse complement motif counts were combined and the motifs with the highest frequency difference between the highest and lowest recombination intervals were noted. Following Cirulli *et al.* (2007), the two windows (six windows for the broad-scale) used were excluded and the frequency of the subset of motifs was regressed using the remaining windows. Results were corrected for multiple comparisons using a sequential Bonferroni correction. Additionally, we analyzed the human motif CCNCCNTNNCCNC (Myers et al. 2008; Stevison et al. 2010) and the *D. melanogaster* motif GTGGAAA (Miller et al. 2012) using the approach described above in the Motif Occurrence section above.

Human comparison

We obtained human recombination data from Kong et al. (2002)(Kong et al. 2002) and genome sequence from a Finnish population, a part of the 1000 genomes project(Altshuler et al. 2010). As above, the sequence was partitioned into intervals of known recombination across human chromosome 1 (used because of its large size). Using the same EMBOSS script "dreg," motif frequency of the 13-mer degenerate motif CCNCCNTNNCCNC (Myers et al. 2008; Baudat et al. 2010) was tallied and a regression looking at motif frequency corrected for interval size and recombination rate was performed. Recombination intervals used for the regression were restricted to the same number of windows and similar recombination range of our *D. pseudoobscura* recombination data (Number of intervals = 140 for both datasets; *D. pseudoobscura* cM range: 0.079-3.97, mean: 0.765, median: 0.487; Human cM range: 0.142-3.11, mean: 0.693, median: 0.505).

Results

Prdm9 homology

Oliver *et al.* (2009) suggested that, although *Prdm9* is essential for fertility in mice, it appears to be absent in *Drosophila melanogaster* and its function in meiosis may be lineage or even species-specific. Previous studies of the PRDM protein family support this conclusion (Fumasoni et al. 2007; Fog et al. 2012). To confirm that this protein is

indeed missing in the Drosophila genus, we BLASTed *Prdm9* and PRDM9 against Drosophila species. Using human , sea urchin, and mouse protein input queries with the BLAST tools blastp and PSI-BLAST, we identified the genes *GA26117*, *CG5245*, *crooked legs*, *meics*, *combgap*, *CG9817*, *Blimp-1*, and *trithorax-related*. All Drosophila proteins identified using BLAST contained zinc finger domains, and *CG9817*, *Blimp-1*, and *trithorax-related* contained SET domains. The maximum amino acid sequence identity ranged between 49% and 38% and the part of the query sequence that was covered ranged between 98% and 73%. Nucleotide input queries using blastn yielded results with a maximum identity between 84% to 97%, but the query only covered between 1% to 13% of the nucleotide sequences of the surveyed genes. These BLAST results, combined with previous data, suggest there is not a *Prdm9* homolog detectable in Drosophila. However, genes identified in this manner, which are proteins that possess SET domains and/or zinc fingers, are candidates that may function similarly to PRDM9.

Table 8: PRDM9 candidate proteins

Gene name (*D. melanogaster* homolog)	**Sequence Motif**	**Protein Domains or notes**
GA23469 (*Blimp-1*)	TGA[TG]ANGGA[GT]AA	SET domain, 4 zinc fingers
GA25755 (*hamlet*)	GAAGATGAGGAANNTGN[CT]NNC	SET domain, 7 zinc fingers
GA26409 (*CG9817*)	NCTTA[AT]NGAGAN[TG]N[TC]	SET domain, 5 zinc fingers
GA25849 (*crooked legs*)	GAC[TG]GNNA[TC]GGGGGGGGGG GGGGGGGGGGGGGGGGGGG	15 zinc fingers

GA26228 (*CG5245*)	[GT][TC]CGNGGGGTNCTNC	6 zinc fingers
GA26117	A[TG][CT]GNNTC[CT]GC[CT][GT][GC]ATNNTNCAN[TC][TG]GANG[TC]GA[TC]	11 zinc fingers
GA21024 (*combgap*)	NN[CT][TG][TC]NN[CT]TNACGNGNGA[TG]G[TC][TG]G[TC][TG]N[TC][TG]G[TC]	10 zinc fingers
GA18168 (*trade embargo*)	TGGNANGCCG[CG]ACNT	5 zinc fingers; meiotic protein
GA21437 (*teflon*)	GNGGNNG[TC][TC]	3 zinc fingers; meiotic protein
GA17308 (*grauzone*)	NANGNN[TG][TC]NNACG[TC]C[TG][TC]GN[TC]NGNC	8 zinc fingers; meiotic protein

Candidate protein analysis

To identify if any proteins function in Drosophila recombination in a similar manner to PRDM9 in humans, we selected a subset of *D. pseudoobscura* Cys_2His_2 zinc finger proteins as candidates. In addition to those Cys_2His_2 zinc finger proteins identified using BLAST above, annotated SET domain proteins and proteins involved in meiotic recombination in Drosophila were targeted (Lyne et al. 2007) (**Table 8**). Cys_2His_2 zinc finger DNA binding residues are determined by amino acids at positions -1, 3, and 6 in relation to the start of the alpha helix (Kaplan et al. 2005). We recorded the amino acid binding residues for all BLAST, SET domain, and meiotic protein candidates, and obtained the predicted nucleotide targets using the approach taken to identify the

binding preferences of PRDM9 (Baudat et al. 2010). Once a consensus motif was established for each protein, a 3 base pair sliding window of 9 base pairs was used for each motif, as the binding length for a protein with N fingers is 3N and not all zinc fingers may be used in binding. Motif occurrence was then analyzed using two *D. pseudoobscura* sequence and recombination datasets: a "superfine-scale" recombination map and a "broad-scale" recombination map. The superfine-scale dataset surveys recombination in 16 intervals of approximately 20 kb in size over 3 Mb of chromosome 2. This dataset was constructed using over 10,000 individuals and contains 219 observed crossover events. The broad-scale dataset estimates recombination in 140 intervals of approximately 180 kb in size across all of chromosome 2 (30 Mb) (McGaugh et al. 2012). This dataset was constructed by genotyping approximately 1400 individuals at 384 markers across the genome and captured 1344 crossover events.

Motif frequency was regressed with recombination rate, and after correcting for multiple comparisons, no motifs were significantly associated with recombination at the superfine scale, and three sequence motifs were significantly associated with recombination at the broad scale. These proteins are *crooked legs* ($p=0.003$, $r=0.331$), which functions in lateral inhibition, cell adhesion, and negative regulation of transcription; GA26117 ($p=0.019$, $r=-0.296$), of unknown biological function; and *combgap* ($p=0.0253$, $r=-0.290$), which functions in imaginal disc-derived wing morphogenesis. These results, while significant, are not particularly compelling due to the high repeat content in the

crooked legs motif, and the highly degenerate nature of the other two motifs (**Table 9**). Of note, *trade embargo* was not significantly associated with recombination at either scale, providing support that it may not bind to discrete foci (Lake et al. 2011).

Table 9: Proteins with motifs significantly associated with recombination

Gene (*D. melanogaster* homolog)	**Sequence Motif**	**p, r (GC content corrected p, r)**	**Differences between zinc fingers in comparison to *D. pseudoobscura***
GA25849 (crooked legs)	GGGGGGGGG	Broad: 0.0034, 0.3307 (0.0003, 0.3321) Superfine: 0.9106, -0.0306 (0.8514, 0.1564)	No differences in mel, per, mir
GA26117	TNNTNCAN[TC]	Broad: 0.01923, -0.2963 (0.0015, 0.3013) Superfine: 0.3510, -0.2497 (0.3050, 0.4086)	NA
	AN[TC][TG]GANG[TC]	Broad: 0.0272, -0.2882 (0.0012, 0.3050) Superfine: 0.9159, 0.0287 (0.5814, 0.2829)	
GA21024 (combgap)	[TG][TC]NN[CT]TNAC	Broad: 0.0253, -0.2902 (0.0017, 0.2979) Superfine: 0.3241, 0.2635 (0.5318, 0.3043)	No differences in mel, mir, or per
GA14083 (CG1663)	NTTACTNT[CT]	Broad: 0.3370, -0.2893 (0.0025, 0.2894) Superfine: 0.005, 0.6644 (0.0227, 0.6644)	7 mismatches in mel, no differences in per or mir
GA11272 (CG11906)	NTTGNG[TG]NTATT	Broad: >1.0, -0.1061 (0.3842, 0.1178) Superfine: 0.0080,0.6362 (0.0343, 0.6362)	3 mismatches in mel, no differences in mir, 1 extra finger in per
GA18222 (CG4496)	GAT[TG]CNGAT	Broad: >1.0, 0.0389 (0.7787, 0.0604) Superfine: 0.0136,0.6019 (0.0530, 0.6030)	2 mismatches in mel, no differences in per or mir
GA11481 (pebbled)	GTTGCTNGN	Broad: >1.0, 0.0248 (0.8067, 0.0560) Superfine: 0.0165,0.5885 (0.0573, 0.5965)	Unknown in mel, no differences in mir, 6 fewer fingers in per
GA18373 (CG4707)	[TG]ATNCGGTT	Broad: >1.0, -0.1866 (0.0571, 0.2023) Superfine: 0.0259,0.5542 (0.0918, 0.5545)	2 mismatches in mel, no differences in per or mir

GA15299 (CG2202)	N[GA]GGGGGGG	Broad: <0.0001, 0.4723 (<0.0001, 0.4702) Superfine: 0.8288, -0.0588 (0.8399, 0.1627)	1 mismatch in mel, 3 fewer fingers in per, no differences in mir
GA21173 (su(Hw))	[CA][CT]TNAG[GC]T	Broad: <0.0001,-0.4444 (<0.0001, 0.4488) Superfine: 0.2679, 0.2946 (0.8471, 0.1588)	1 mismatch in mel, 1 fewer finger in per, missing sequence in mir
GA12131 (zfh1)	GTTANNNTN	Broad: 0.0050,-0.3676 (<0.0001, 0.3678) Superfine: 0.7078, 0.1017 (0.8518, 0.1561)	1 extra finger in mel no differences in per, missing sequence in mir
GA22134 (CG9932)	NNTANN[GC][TC]N	Broad: 0.0083,-0.3592 (<0.0001, 0.3621) Superfine: 0.4855, -0.1881 (0.5491, 0.2968)	3 mismatches in mel, no differences in per or mir
GA14502 (Oaz)	[GC]TTANNGNN	Broad: 0.0166,-0.3474 (0.0001, 0.3535) Superfine: 0.1056, -0.4197 (0.3145, 0.4037)	Unknown in mel, no differences in per or mir
	TNTT[CA][GA]G	Broad: 0.0234,-0.3413 (0.0002, 0.3413) Superfine: 0.5369, -0.1668 (0.7461, 0.2099)	
GA20521 (CG7691)	NACNTN	Broad: 0.0219,-0.3424 (0.0001, 0.3481) Superfine: 0.1618, -0.3672 (0.5137, 0.3121)	No differences in mel, per, or mir
GA11205 (charlatan)	NNTN[TG]GG[AT]C	Broad: 0.0328, -0.3351 (0.0002, 0.3401) Superfine: 0.4588, -0.1995 (0.8200, 0.1734)	1 mismatch in mel, no differences in per or mir
GA11270 (CG11902)	[CA]ATN[TG]G[GC]A[CT]	Broad: 0.0389,-0.332 (0.0003, 0.3374) Superfine: 0.1401, 0.3857 (0.7821, 0.1926)	6 extra fingers in mel and 5 mismatches, no differences in per or mir
GA15842 (CG30431)	NNTATT[GC]NG	Broad: 0.042,-0.3305 (0.0003, 0.3352) Superfine: 0.9141, 0.0293 (0.7495, 0.2083)	No differences in mel, per, or mir

A comprehensive search for a zinc finger binding sequence motif

To determine if *any D. pseudoobscura* Cys_2His_2 zinc finger proteins are associated with recombination, the amino acid binding residues of all *D. pseudoobscura* zinc finger proteins were recorded, and predicted nucleotide targets were generated using the program enoLOGOS (Workman et al. 2005). Sequence motifs for 186 *D. pseudoobscura* proteins were identified. Again, a 3 base pair sliding window of 9 base pairs was used for each motif, and motif occurrence was determined and analyzed using both the superfine- and broad-scale *D. pseudoobscura* sequence and recombination datasets (McGaugh et al. 2012). The superfine scale recombination dataset was analyzed using all nucleotide sequence and strictly intergenic sequence, while the broad-scale dataset was analyzed using all nucleotide sequence. Intergenic sequence was used for the superfine scale recombination dataset because recombination is known to commonly initiate in intergenic regions in yeast (Petes 2001; Pan et al. 2011).

After correcting for multiple comparisons, no Cys_2His_2 zinc finger protein analyzed at the superfine scale was significant. This is complicated by the fact that there are 640 comparisons and only 16 windows of known recombination, so power to detect significance is low. At the broad scale, ten proteins were significantly associated with recombination after correction for multiple comparisons (**Table 9**). Only one protein, GA15299 (CG2202), was positively associated with recombination. The remaining nine proteins were negatively associated with recombination, which may be expected for a

protein like suppressor of Hairy wing (*su(Hw)*) that functions in the negative regulation of transcription and negative regulation of chromatin silencing, but contradicts the expectation from PRDM9. Furthermore, there was no overlap in zinc finger motifs between the fine scale and broad scale analyses, which casts doubt on the detected associations.

Differences in motif occurrences do not account for changes in recombination landscapes between closely related species

PRDM9 is known to be undergoing rapid positive selection, changing both the number of zinc fingers present and the DNA-binding amino acid residues at positions -1, 3, and 6 (Oliver et al. 2009; Ponting 2011). To determine Cys_2His_2 zinc finger proteins changing rapidly across the Drosophila lineage, we compared number of zinc fingers present and number of changes in binding residues for each Cys_2His_2 protein in *D. melanogaster* (55 mya), *D. miranda* (3 mya), and *D. persimilis* (0.5-1 mya) to *D. pseudoobscura*.

Between *D. pseudoobscura* and *D. miranda*, a large majority of Cys_2His_2 zinc finger protein binding residues are conserved. To identify any changes in recombination rate associated with change in binding of Cys_2His_2 zinc finger proteins, we generated new sequence motif predictions for proteins with mismatches between *D. miranda* and *D. pseudoobscura*. Utilizing a *D. miranda* broad-scale recombination map directly comparable to the broad-scale *D. pseudoobscura* recombination map, we then compared the

association between *D. miranda* binding motifs and *D. miranda* recombination to *D. miranda* binding motifs and *D. pseudoobscura* recombination and then repeated with comparing *D. pseudoobscura* motifs to *D. pseudoobscura* and *D. miranda* recombination. If a protein is involved in recombination, we expect to see a stronger correlation between binding motif and recombination within species than between species (**Figure 10**). Comparing all associations with an un-corrected p-value <0.05, no protein was consistently more strongly correlated with recombination within species than between species. Therefore, we conclude that no differences associated with zinc finger binding are responsible for recombination rate changes between these species of Drosophila.

	D. miranda **Recombination & Sequence**	***D. pseudoobscura*** **Recombination & Sequence**
D. miranda **Predicted motif**	correlation coefficient (r) >	correlation coefficient (r)
D. pseudoobscura **Predicted motif**	correlation coefficient (r) <	correlation coefficient (r)

Figure 10: Predictions to test if changes in DNA binding motifs between species account for variation in recombination rate between species

Sequence motifs

To identify overrepresented sequence motifs without an identified association with zinc fingers, all possible 6 base pair motifs were analyzed for frequency using the superfine and broad scale recombination datasets. Ten motifs with the greatest frequency difference between regions of high and low recombination intervals were selected and regressed with recombination rate. At the broad scale, the motifs AATAAA ($p=0.0397$, $r=-0.178$) and CTGCTG ($p=0.0539$, $r=-0.1669$) were weakly, negatively associated with recombination and the motifs CTCTCT ($p=0.0115$, $r=0.0115$) and TCTCTC ($p=0.0126$, $r=0.2149$) were weakly, positively associated with recombination. At the superfine scale, the motifs AAATTT ($p=0.0717$, $r=0.4954$) and ACAAAT ($p=0.0594$, $r=0.5151$) were weakly, positively associated with recombination.

Previous studies in Drosophila found associations between local recombination rates and the human *Prdm9* motif CCNCCNTNNCCNC (Stevison et al. 2010), and the *D. melanogaster* motif GTGGAAA (Miller et al. 2012). In this study, neither of these previously described motifs were significantly associated with recombination rate variation in *D. pseudoobscura* at either scale, although this lack of association is not unexpected as these motif associations were detected previously in different species.

A validation of our approach using human recombination data

To assess if one can detect an association between a sequence motif and recombination rate using relatively coarse recombination rate estimates, we utilized recombination data from an Icelandic population that empirically surveyed genome wide recombination in 869 individuals (average window size: 650kb). We restricted the dataset to a subset of chromosome one with a comparable recombination range to *D. pseudoobscura*. A regression between the frequency of the human *Prdm9* motif CCNCCNTNNCCNC and recombination rate was positive and statistically significant ($p=0.0004$, $r=0.3$), thus demonstrating sequence motif signals can be detected in humans with broad-scale recombination data comparable to that used in the Drosophila studies.

Discussion

Our attempts to identify a PRDM9-like protein involved in meiotic recombination initiation in Drosophila yielded negative results. Generating predicted nucleotide sequence motifs from Cys_2His_2 zinc finger proteins and regressing their frequency with estimated recombination in *D. pseudoobscura* produced a handful of recombination associated sequence motifs, but the biological relevance of these associations remains uncertain. Furthermore, changes in the binding motifs between species do not appear to account for variation in the recombination landscape. Our

results could be complicated by the approach taken, or alternatively, we suggest these findings could be explained by the existence of a different recombination initiation system in Drosophila.

Approach

Our results could be due in part to the scale at which recombination was estimated in *D. pseudoobscura*. While the superfine (20kb) and broad scale (180kb) recombination datasets used represent one of the most comprehensive recombination maps outside human, mouse, and yeast, the datasets might still lack the resolution needed to determine sequence motifs associated with recombination. Successful work with sequence motifs in yeast and human recombination has been analyzed at a scale <1-2 kb (Myers et al. 2005; Myers et al. 2008; Steiner et al. 2009; Pan et al. 2011; Steiner et al. 2011), although we were able to detect a strong association here between the frequency of the human *Prdm9* motif and human recombination rate using broader (~650kb intervals), comparable in recombination to what we studied, thereby validating our approach. Furthermore, recombination associated motifs have been identified at scales ranging from 220kb to 5 Mb in other organisms (Shifman et al. 2006; Groenen et al. 2009; Backstrom et al. 2010; Stevison et al. 2010; Wong et al. 2010).

Additionally, there is an inherent limitation in one of the bioinformatic approaches utilized here, in the ability of currently developed programs to accurately

identify DNA binding motifs of zinc fingers. While algorithms have improved over the years, it is impossible to be certain that identified motifs are "correct." Hence some motifs predicted in this manner could be biologically irrelevant. We attempted to address this problem in two ways. First, for BLAST, SET domain, and meiotic candidate zinc finger proteins, we followed a proven motif prediction protocol: the methods utilized to identify the binding nucleotides for PRDM9 (Baudat et al. 2010).Because of its success in determining the PRDM9 associated sequence motif, we can be somewhat more confident in concluding that our motif predictions for these proteins are correct, and therefore, that no Drosophila candidate proteins we tested are associated with recombination. Second, we used an unbiased approach to identify all motifs of six base pairs in length and to test their association with recombination, although even this approach is imperfect since it is not possible to search for degenerate motifs of all possible lengths. Despite these accommodations, it remains possible that Cys_2His_2 zinc finger protein associated sequence motifs do play a role in Drosophila meiotic recombination, but that it is beyond the scope of the technology to detect them at this point in time.

A different recombination initiation system in Drosophila?

Alternatively, it is possible and likely that other factors play a major role in the determination of recombination in Drosophila. Historically, it has been thought that

Drosophila do not have the 1-2 kb hotspots characteristic of yeast, humans, and mice (Aquadro et al. 2001; Nachman 2002; Hey 2004; Coop et al. 2007). This is supported by the lack of apparent hotspots of intragenic recombination in *rosy* (Clark et al. 1988; Hilliker et al. 1991; Radford et al. 2007) and in *white-echinus* (Singh et al. 2009), and the lack of linkage disequilibrium among nearby nucleotides as compared to humans (Aquadro et al. 2001; Ardlie et al. 2002; Nachman 2002). The human *Prdm9* recombination initiation model is based on the specific targeting and binding of the PRDM9 protein to a sequence motif, enriched in recombination hotspots (although this model may need some refining, see below). If Drosophila lack such recombination hotspots, this evidence supports Drosophila lacking a recombination initiation system that functions in a sequence specific binding function like *Prdm9* in humans, although obviously cannot rule out a sequence binding independent function.

Additionally, Drosophila recombination is known to differ from other organisms (Hawley et al. 1992; Roeder 1997; Keeney 2001; Hawley et al. 2002; McKim et al. 2002; Page et al. 2003). First, homologous chromosome pairing and synapsis proceed normally in the absence of double strand breaks. Second, the synaptonemal complex does not require SPO11 to form and functions in the initiation of recombination and the facilitation of DSBs. Third, Drosophila are missing some genes known to be crucial in recombination in other organisms (McKim et al. 2002). With these known differences in meiotic proteins, and apparent differences in the initiation of recombination, this

evidence is supportive of Drosophila possessing a different recombination initiation process than humans.

Furthermore, *Prdm9* is missing or altered in many organisms (Fumasoni et al. 2007; Oliver et al. 2009; Ponting 2011), necessitating the existence of alternative recombination initiation systems. The PRDM family is absent in plants and fungi, and is quite small in other taxa, with only two genes in nematodes and three genes in arthropods (*D. melanogaster*: *CG5249* –PRDM1, *CG9817* – PRDM5, and *hamlet*). While PRDM9 functions in meiotic recombination in both mouse and human, it seems as if this function is lineage specific. *Prdm9* is non-functional in canines (Munoz-Fuentes et al. 2011; Axelsson et al. 2012), and is missing all zinc fingers in the marsupial *Monodelphis domestica* (Ponting 2011), so even amongst mammals, recombination initiation may vary.

Finally, the PRDM9 story is made more complex by a general lack of understanding of the *in vivo* function of PRDM9. In humans, although the PRDM9 motif is only detected in a proportion of hotspots, data suggest that PRDM9 influences hotspot activity even at hotspots in which the motif is absent(Berg et al. 2010; Berg et al. 2011; Segurel et al. 2011). While there is *in vitro* evidence that the zinc fingers of PRDM9 do bind to the motif, this suggests PRDM9 interacts with hotspots genome wide in a more complex and subtle way than first expected. Furthermore, the predominant human sequence motif is neither necessary nor sufficient to drive hotspot activity in humans, with the motif represented approximately 290,000 times in the genome and only about

50,000 detectable hotspots(Segurel et al. 2011). In chimpanzees, there is extensive variation in the PRDM9 protein, and little evidence of any sequence motifs enriched in hotspots(Auton et al. 2012). Researchers suggest the most plausible explanation for this observation is that PRDM9 may still play the same roles in chimpanzee as it does in mouse and human, but the PRDM9 alleles may bind to a much greater variety of sequence motifs than seen in human. This implies that other factors, like chromatin state, play a more dominant role in the hotspot localization. Taken in this context, our data could suggest that there is a PRDM9-like protein in Drosophila, but it either binds a wide repertoire of sequence motifs, or functions in a sequence-specific-binding independent manner.

Regardless of the model, given recent observations that PRDM9 influences more human recombination hotspots than previously thought, and possibly all hotspots (Segurel et al. 2011), it is quite remarkable that a single protein rapidly evolved to become responsible for all recombination in the human lineage. Recombination is an essential mechanistic and evolutionary process, so *Prdm9* poses an intriguing step in the evolution of meiosis. However, *Prdm9* appears to be only a piece of the puzzle when looking at recombination across taxa. Evidence from Drosophila and other organisms suggests that *Prdm9* is not the quintessential element defining meiotic recombination; instead, there remain many mysteries to explore.

4. No detectable effect of DNA methyltransferase DNMT2 on meiotic recombination

Caiti S. Smukowski Heil

Introduction

Epigenetics has long been predicted to play a role in the initiation of meiotic recombination. Observations of variation in recombination rate within and between individuals, sexes, populations, and across the genome (such as euchromatin versus heterochromatin) suggest a role beyond DNA sequence in determining locations of recombination events (Lichten et al. 1995; Barthes 2011). An association between open chromatin formation and double strand breaks, the first step in the initiation of recombination, has been identified in yeast, dog, and several plants (Berchowitz et al. 2009; Pan et al. 2011; Auton et al. 2013; Choi et al. 2013; Hellsten et al. 2013), and the histone H3K4 methyltransferase PRDM9 influences the distribution of recombination sites in human and mouse (Buard et al. 2009; Baudat et al. 2010; Berg et al. 2011; Grey et al. 2011; Brick et al. 2012; Acquaviva et al. 2013; Sommermeyer et al. 2013). However, the possible roles of specific epigenetic marks apart from H3K4me3 are less understood. For example, the relation between recombination initiation and DNA methylation, the best characterized epigenetic factor, appears complex and remains relatively unexplored.

DNA methylation describes the transferring of a methyl group (CH_3) to the 5^{th} position of a cytosine residue, typically at CpG sites and repeat elements (Robertson et

al. 2000). The reaction is catalyzed by a family of conserved proteins known as DNA methyltransferases: DNMT1, the maintenance methyltransferase, ensures proper inheritance of methylation patterning after replication in somatic cells; DNMT3s (3A, 3B, 3L), the de novo methyltransferases, establish DNA methylation patterns during embryogenesis; and DNMT2, an enigmatic methyltransferase with conserved catalytic motifs, has a historically disputed function (Yoder et al. 1997; Dong et al. 2001; Schaefer et al. 2010; Barthes 2011; Krauss et al. 2011). These genes function within an ancient regulatory mechanism shared by animals, plants, and fungi , serving in diverse roles often related to repression of gene expression (Feng et al. 2010; Zemach et al. 2010; Zemach et al. 2010; Jurkowski et al. 2011; Nanty et al. 2011).

A link between DNA methylation and recombination was first hypothesized by Rossignol and Fagueron (Rossignol et al. 1994) and Yoder *et al.*(Yoder et al. 1997), in which DNA methylation promotes genome integrity through the suppression of recombination between dispersed repetitive sequence. Such ectopic recombination events result in gross chromosomal rearrangements of deleterious consequence (Lichten et al. 1987; Rouyer et al. 1987; Montgomery et al. 1991; Small et al. 1997; Li et al. 2012), and because the prevalence of repetitive elements in a genome is so high, a system must exist to limit recombination between repeats. Indeed, early evidence showed that repeats were enriched for methylation in plants, fungi, and mammals (Whitkus et al. 1992; Bennetzen et al. 1994; Yoder et al. 1997; Maloisel et al. 1998), and the loss of DNMT3A or

DNMT3L resulted in reactivation of retrotransposons, impaired synapsis of homologous chromosomes, failure to progress through meiotic prophase, and sterility (Bourc'his et al. 2001; Bourc'his et al. 2004; Kaneda et al. 2004).

Direct evidence of a functional link between DNA methylation and recombination is somewhat limited, but a study in the fungus *Ascobolus immersus* showed crossover formation was reduced several hundred fold in an *in vivo* methylated hotspot compared to an unmethylated hotspot (Maloisel et al. 1998). In more recent years, several reports in Arabidopsis paint a more nuanced picture. For example, in the absence of MET1 (the DNMT1 homolog), researchers independently observed a pattern of increased recombination in euchromatin, and decreased recombination in (typically hypermethylated, transposon rich) heterochromatin (Melamed-Bessudo et al. 2012; Mirouze et al. 2012; Yelina et al. 2012). Each study found that total number of crossover events was not different between *met1* mutants and wild type, indicating that the loss of DNA methylation affects the distribution of crossovers, but not their overall number. Looking specifically at wild type Arabidopsis transcription start (TSS) and termination sites (TTS), DNA methylation was decreased in recombination hotspots relative to TSS or TTTs in which recombination was absent (Choi et al. 2013), supporting data from *A. immersus*. In contrast, some indirect evidence in humans found a positive association between recombination rate and DNA methylation (Sigurdsson et al. 2009).

In this study, I explore the complex and perhaps contradictory role of DNA methylation in the determination of recombination events in the model system *Drosophila melanogaster*. *D. melanogaster* possesses the DNA methyltransferase DNMT2 and a methyl binding domain protein, MBD2/3, which typically binds to methylated DNA and recruits chromatin remodeling complexes (Tweedie et al. 1999; Roder et al. 2000; Ballestar et al. 2001; Marhold et al. 2004). Transcripts of *Dnmt2* were particularly enriched during early stages of embryonic development; expression in adult flies was limited to female ovaries and there was no activity in male testes (Lyko et al. 2000), perhaps indicating a role in recombination, which is female-specific in Drosophila. Furthermore, like MBD2 knockout mice, Drosophila null mutants of MBD2/3 were viable and fertile, but revealed chromosome segregation defects (Lyko et al. 2006).

DNA methylation was purportedly detected at low levels, ranging from 0.21-0.4% across the Drosophila genus (Tweedie et al. 1999; Gowher et al. 2000; Lyko et al. 2003; Marhold et al. 2004; Salzberg et al. 2004; D'Avila et al. 2010). One of its functions in Drosophila was thought to be retrotransposon silencing and stabilization of repeats, similar to its role in vertebrates and plants (Salzberg et al. 2004; Phalke et al. 2009; Krauss et al. 2011). Although methylation levels may be low (or maybe even virtually absent, see Raddatz et al. 2013), it appears that DNMT2 is functional in a capacity similar to its role in other taxa. For example, in somatic cells, the loss of *Dnmt2* eliminates H4K20 trimethylation at retrotransposons and impairs maintenance of retrotransposon

silencing (Phalke et al. 2009), and Su(var)3-9 (H3-K9 specific histone methyltransferase) mutants, normally associated with inactive regions in the genome such as constitutive and facultative heterochromatin and genome stability, showed dramatic reduction or complete loss of DNA methylation (Lyko et al. 2003).

Therefore, to determine if there is a detectable effect of DNA methylation and/ or the DNA methyltransferase *Dnmt2* on the distribution and frequency of recombination rate in Drosophila, I assayed recombination at two euchromatic regions on the X chromosome in *Dnmt2* -/- and control *D. melanogaster*. I did not detect any change in recombination rate or distribution in the absence of *Dnmt2*, and conclude that other factors are determining sites of recombination events in Drosophila.

Materials & Methods

Stocks and crossing scheme

For all crosses, virgin flies were collected, separated by sex, and aged for 7 days. The crossing scheme (**Figure 11**) consisted of: (1) crossing a *D. melanogaster Dnmt2* p-element excision line Dnmt2^{99}(Schaefer et al. 2010) to wild type *D. melanogaster* Zim29 to generate variability to score recombination events. (2) F_1 females were crossed to a *D. melanogaster* chromosome 2L deficiency line over a balancer, Df(2L)BSC826/SM6a (#27900 Bloomington Stock Center, Bloomington, IN) . F_2 females were collected, and females carrying the SM6a balancer were identified by the Cy phenotype and discarded.

(3) The remaining F_2 females (bearing 0-1 functional copies of *Dnmt2*) were crossed to wild type males (Zim29) in single pair crosses, allowed to lay eggs, and subsequently genotyped after larvae appeared. Undesired genotypes (see below for genotyping methods) were discarded and F3 progeny were collected from remaining vials. Recombination was assayed in these individuals.

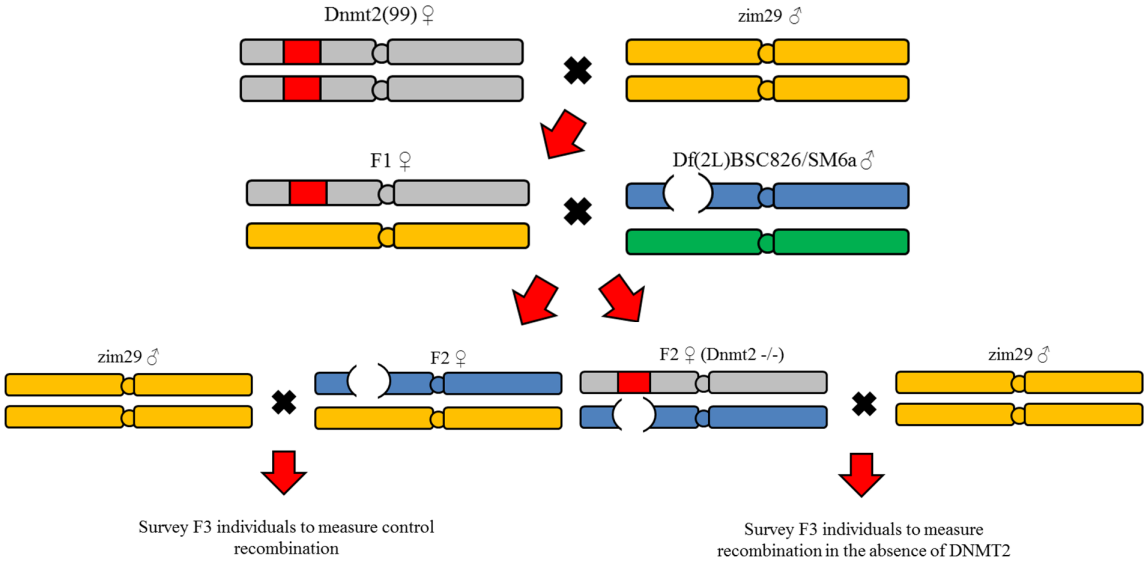

Figure 11: DNA methylation crossing scheme

Reverse transcription

To ensure that individuals of $Dnmt2^{99}$/ Df(2L)BSC826 genotype were *Dnmt2*-/-, reverse transcription PCR was completed. For each line ($Dnmt2^{99}$, Df(2L)BSC826/SM6a, zim29, $Dnmt2^{99}$/ Df(2L)BSC826), I prepped RNA from approximately 25 pooled flies of varying life stages using the Qiagen RNeasy kit, Qiagen QIAShredder kit, and Qiagen

DNase kit. For each line, we used genomic DNA, H_2O, and a negative control reverse transcription product as controls. For the reverse transcription reaction, the following recipe was used: 2µl 10X PCR buffer (15mM MgCl2), 1.5µl 50mM MgCl2, 0.8µl 25mM dNTPs, 1µl 50µM experimental R primer, 1µl 50µM control intron R primer, 0.5µl 40U/µl RNase inhibitor, 0.1µl 200U/µl MMLV reverse transcriptase, 9.1µl DEPC H2O, and 4µl RNA mix + H2O to get to 500ng RNA. The reverse transcription program consisted of: 15 minutes at 42° C followed by five minutes at 99° C. Then a PCR was completed with the following recipe: 2.5µl 10X PCR buffer (15mM MgCl2), 1.5µl 2mM dNTPs, 1.25µl, 10µM experimental F primer, 1.25µl 10µM experimental R primer, 0.3µl DNA polymerase, 16.2µl H_2O, and 2µl reverse transcription reaction product. The PCR program consisted of an initial denaturing step of 95ºC (60 sec), three touch-down cycles of 94ºC (30 sec)- 56ºC (30 sec)- 72ºC (45 sec) each, three touch-down cycles of 94ºC (30 sec)- 53ºC (30 sec)- 72ºC (45 sec) each, followed by 33 main cycles of 94ºC (30 sec)- 50ºC (30 sec)- 72ºC (45 sec) each. We used the *trade embargo* (TREM) gene as a control with the following primers: Forward: CAGTAAGTGTGAATCCTGCTTGGTTTGC; Reverse: GCATGTCCATAATGTGCTGATGGGATC. The primers used for flanking the *Dnmt2* gen intron were: Forward: GGTCTTAGAACTATTTAGTGGCATTGGCG; Reverse: TAATTGTGCGCATAAACCGCATTGGC.

Scoring recombination

Flies were collected in 96 well plates and frozen at -20° C. DNA was extracted following the protocol of Gloor and Engels (Gloor 1992), using 49.5ul squish buffer (10mM Tris-HCl (pH 8.2), 1mM EDTA, 25mM NaCl) + 0.5ul proteinase-K. A zirconium bead was placed in each well, and plates were then shaken using a Qiagen TissueLyser II for 45 seconds. The DNA preps were then incubated at 37°C for 30 minutes and 95°C for 2 minutes in a thermal cycler. The PCR recipe consisted of 0.5 uL of forward primer +M13 tag (CACGACGTTGTAAAACGAC added to 5' end of forward primer) , 0.5 uL of reverse primer, 0.4 uL of 700IRD or 800IRD-labeled M13 tag, 1.3 uL of MgCl2, 10 X buffer, 1 uL 2 mM dNTPs, and 0.2 ul of Taq polymerase in a 10 uL reaction volume. The same PCR program was used as above. Products were visualized on a polyacrylamide gel using a LICOR 4300.

To identify crosses with desired genotypes, F_2 females were genotyped at three loci on the X chromosome and two loci on Chromosome 3(**Table 10**) to ensure heterozygosity (Df(2L)BSC826/zim29) across the region of interest, and at one locus on chromosome 2L (**Table 10**) to identify if the genotype was Dnmt2^{99}/ Df(2L)BSC826 (*Dnmt2*-/-) or Df(2L)BSC826/zim29 (control). Vials from parents of the desired genotypes were kept and the F_3 progeny were collected, all other vials and their progeny were discarded.

Recombination was scored in F_3 progeny by genotyping at the same three markers on the X chromosome and two markers on the 3rd chromosome (**Table 10**). The

markers on the X chromosome delineate 2 regions, one of lower recombination (~1.8 cM/Mb) and one of higher recombination (~3.7 cM/Mb). A recombinant was called when an individual fly's genotype changed from heterozygous to homozygous or vice versa for females, and when the fly's genotype changed between the possible allele combinations for the males. In total, 1,536 F3 control progeny and 1,177 F_3 experimental progeny were scored for the euchromatic regions.

Table 10: Markers used to assay recombination

Marker	**Primer name**	**Physical distance/genetic distance between markers**	**Sequence (F/R)**
Marker X1	DMELX_494471F DMELX_494621R	Marker X1-Marker X2: 2.75 Mb/5 cM	CGAGCGCTGTCTATTGCGTTC TCATTTCAATTCCGATTTGGAGTCGGC
Marker X2	DMELX_3240050F DMELX_3240200R	Marker X2-Marker X3: 2.68 Mb/10 cM	GGAAACAGTGTTATTGCCTACACATGG AAC CTTGGCCAAGTTGCACATGAGATAC
Marker X3	DMELX_5922532F DMELX_5922673R	Marker X1-X3: 5.43 Mb/15 cM	GGATCGTTGCAGATCGGATAGAACTC CCGTCTCAAATTGATGGACGCCTAT
Marker Chr2	DMEL2L_12024260F DMEL2L_12024434R	NA	CGTCACATTCCATTGAACGACTTTCGG CAAAACTGGCTCCAAACGTCCGTG

Statistics

Recombination fractions between experimental and control individuals were compared using an unpaired *t*-test (GraphPad Software, Inc. La Jolla, CA). A power

analysis was completed using the "pwr" package in R (Statistic: (Cohen 1988) R package: Stephane Champely).

Results

Recombination across a euchromatic region

To identify an effect of the DNA methyltransferase gene *Dnmt2* on meiotic recombination, we created a variable *Dnmt2* null fly stock and identified recombinants in two adjacent euchromatin regions on the X chromosome. Region 1 spans assembly positions 494,471 to 3,240,200 bp (Adams et al. 2000) and represents a region of relatively low recombination (2.75 Mb, approximately 5 cM), while Region 2 spans assembly positions 3,240,200 to 5,922,673 bp and represents a region of relatively high recombination (2.68 Mb, 10 cM). We chose this delineation to account for possible changes in the distribution of recombination events in euchromatin in the absence of *Dnmt2*.

In total, 1536 control (*Dnmt2*+/-) individuals and 1177 *Dnmt2*-/- individuals were scored. In Region 1, we identified no significant difference between *Dnmt2* -/- and control individuals (p=0.86; Control: 3.02% recombinant, 1.07 cM/Mb; Experimental: 3.16% recombinant, 1.11 cM/Mb). Results were similar for Region 2 (p=0.84 ;Control: 10.58% recombinant, 3.57 cM/Mb; Experimental: 8.98% recombinant, 3.07 cM/Mb). A

power analysis showed with our sample sizes, we could detect an effect size of d=0.1 (power=0.8, significance level 0.05).

Discussion

Although DNA methylation influences the recombination landscape in fungus *A. immersus* and Arabidopsis, we detected no effect of knocking out the DNA methyltransferase *Dnmt2* on recombination in Drosophila. Our crossover analyses were limited to two genomic regions, but these regions captured both low- and high-recombination areas of euchromatin. As such, DNMT2 appears to have no major effect on rates of recombination in specific regions of the genome, and DNA methylation more generally may also not affect recombination in Drosophila.

The presence of DNA methylation in Drosophila and other *Dnmt2*-only systems is a long debated issue (Schaefer et al. 2010; Raddatz et al. 2013). While *Dnmt2* shows strong sequence and structural conservation to established methyltransferases, the enzymatic activity was found to be much weaker (Okano et al. 1998; Dong et al. 2001). Various studies reporting DNA methylation in Drosophila (Tweedie et al. 1999; Gowher et al. 2000; Lyko et al. 2000; Lyko 2001; Lyko et al. 2003; Lyko et al. 2003; Marhold et al. 2004; Marhold et al. 2004; Salzberg et al. 2004; Phalke et al. 2009; Krauss et al. 2011) may have been confounded by contaminations from other organisms, detection limits, low antibody specificity, and/or false positives (Zemach et al. 2010; Raddatz et al. 2013).

More advanced bisulfite sequencing with thorough controls revealed no detectable levels of DNA methylation in Drosophila (Raddatz et al. 2013).

Our finding that there is no effect of the gene *Dnmt2* on meiotic recombination supports this recent data, and points to other epigenetic mechanisms directing recombination in Drosophila. Indeed, there is some evidence that histone modifications in Drosophila may mimic the role of DNA methylation in transcriptional processes in other invertebrates (Cedar et al. 2009; Chodavarapu et al. 2010; Nanty et al. 2011; Hunt et al. 2013). Additionally, although DNA methylation was reportedly involved in specific transposons in Drosophila (Phalke et al. 2009), the small RNA Piwi-piRNA pathway is known to be the main genome defense system against repetitive elements in the germline(Aravin et al. 2007; Brennecke et al. 2007; Blumenstiel 2011).

Irrespective of its lack of clear role in recombination, new data show that *Dnmt2* may still play a role in the control of RNA viruses in an unanticipated way: a necessity of *Dnmt2* for acute immune responses after RNA virus infection and the presence of RNA virus genomes in *Dnmt2* protein complexes points to a role in innate immune response, perhaps via RNA methylation (Durdevic et al. 2013; Durdevic et al. 2013; Durdevic et al. 2013). Goll *et al.* (2006) first showed that *Dnmt2* methylates cytosine 38 of $tRNA^{Asp}$, and while at that time it was thought to also methylate DNA, RNA may perhaps be its only genuine substrate (with possible "star activity:" low enzymatic activity with relaxed substrate specificity on DNA substrates). It's likely that the tRNA

methylation mechanism is conserved between DNA methyltransferases, and phylogenetic analyses support the idea that DNMT2 is a non-canonical member of the DNA methyltransferase family, not an RNA methyltransferase (Jeltsch et al. 2006; Jurkowski et al. 2008; Jurkowski et al. 2011). While Dnmt2's biological function is still being explored, observations in Drosophila and other organisms indicate that *Dnmt2* mediates assorted forms of "RNA stress," triggered by events such as an excess of retro-transposons or RNA viruses (Lin et al. 2005; Phalke et al. 2009; Schaefer et al. 2010; Becker et al. 2012; Tuorto et al. 2012; Durdevic et al. 2013; Durdevic et al. 2013). Finally, new evidence in mice and Drosophila has shown that *Dnmt2* is required for trans-generational RNA-mediated epigenetic heredity via paramutation, maintaining *Dnmt2*'s function as an epigenetic factor (Kiani et al. 2013).

Clearly *Dnmt2* research has experienced a tumultuous twenty years, with some long debated questions answered (e.g. (Raddatz et al. 2013)), and many new ones posed (e.g. (Durdevic et al. 2013; Kiani et al. 2013)). Whether DNA methylation influences the recombination landscape in organisms besides the fungus *A. immersus* and Arabidopsis is one question that remains to be elucidated, but I conclude based on the available results that it has no role in Drosophila. While both recombination and the *Dnmt2* phenotype are known to be affected by environment and stress (Stevison 2009; Durdevic et al. 2013), and though it remains possible that *Dnmt2* methylates DNA in stress conditions or via unknown environmental cues, this seems unlikely (Krauss et al. 2011;

Raddatz et al. 2013). Future investigation of such a role should be pursued in organisms possessing DNMT1 and DNMT3.

5. The fine scale recombination landscape in two species of Drosophila

Caiti S. Smukowski Heil, Edwin Iversen, and Mohamed Noor

Introduction

Recombination is a key evolutionary process, affecting adaptation, speciation, sex chromosome evolution, and the distribution of genomic features. Sexual reproduction is thought to persist in part because homologous meiotic recombination breaks down linkage and permits sites to behave independently, theoretically enhancing the efficacy of selection, allowing the purging of deleterious variants, and accelerating adaptation (Smith et al. 1974; Kondrashov 1988; Otto et al. 2001; Otto et al. 2002; Haddrill et al. 2007; Barton 2010). Theoretical predictions are supported by observations of sex chromosomes, where deleterious mutations accumulate and cause the genome to deteriorate in regions of no recombination (Charlesworth et al. 2000; Bachtrog et al. 2002; Bachtrog 2003), and experimental studies, in which organisms with recombination experience greater rates of adaptation when selection is strong (Colegrave 2002; Goddard et al. 2005; Morran et al. 2009; Webster et al. 2012).

Recombination both influences the efficacy of selection, and can be selected upon (Chinnici 1971; Charlesworth et al. 1985). Which selective pressures act on recombination is dependent upon the genomic scale examined. At broad scales, one of the most striking general patterns of recombination is the necessity for at least one

crossover per chromosome to ensure proper segregation of chromosomes during meiosis (Roeder 1997; Sapienza 2001 ; Dumont et al. 2008; Fledel-Alon et al. 2009). The physical link created during a crossover event provides tension that is needed for the spindle to pull the correct chromatids to the poles (Roeder 1997); breakdown of this process can cause aneuploidy and cell death. This mandates a minimum level of recombination, as a crossover rate that is too low is likely to be highly deleterious.

Contrarily, too much recombination breaks apart advantageous allele combinations and threatens genomic integrity (Kauppi et al. 2004; Coop et al. 2007). Ectopic recombination, or recombination events that occur between non-homologous regions, such as repetitive sequence, can result in gross, deleterious chromosomal rearrangements and disease (Inoue et al. 2002). Therefore, on a broad scale, recombination is likely constrained between a lower bound dictated by chromosome segregation, and an upper bound determined by maintaining genomic integrity. Studies of recombination rate variation across species support this broad level of constraint: recombination rates on the order of tens of kilobases up to genome wide genetic maps show conservation between closely related species (Ptak et al. 2005; Dumont et al. 2008; Paigen et al. 2008; Backstrom et al. 2010; Meznar et al. 2010; Stevison et al. 2010; Smukowski et al. 2011; McGaugh et al. 2012).

Within these bounds, where a crossover occurs on a chromosome is non-random. Observations across many species, such as yeast, Drosophila, mice, and human reveal

chromosomal domains of high and low recombination, commonly with reduced recombination around centromeres. Evidence in humans suggests that recombination too close to the centromere or telomere elevates rates of non-disjunction (Hassold et al. 2001). Apart from chromosomal domains, recombination is also influenced at intermediate scales by crossover interference. Crossover interference describes the procedure in which a crossover event reduces the probability of a second such event in its proximity, which leads to a more equal distribution of crossovers throughout the genome and likely prevents non-disjunction as well (Chua et al. 1997; Kaback et al. 1999; Broman et al. 2000; Broman et al. 2002; Bishop et al. 2004).

Finally, in most organisms observed, recombination takes place in discrete genomic regions called "hotspots," typically 1-2 kilobases in length in which recombination increases a magnitude fold above the background rate. At fine scales, the recombination landscape is shaped by a phenomenon known as the "hotspot paradox" (Boulton et al. 1997; Jeffreys et al. 2002; Jeffreys et al. 2005). The hotspot paradox describes the persistence of hotspots despite a mechanistic process that results in their loss. More specifically, when a double strand break occurs to initiate a recombination event, the initiating allele is repaired using the non-initiating allele as a template, thereby converting active alleles to inactive alleles, which should lead to the disappearance of hotspots over time. A method to counteract the loss of hotspots was unknown until the recent discovery of the zinc finger histone-methyltransferase, *Prdm9*,

which controls the distribution of recombination in mice and humans (Baudat et al. 2010; Parvanov et al. 2010). The amino acids that determine DNA-binding of *Prdm9* are under rapid, positive selection, and minisatellite shuffling has added, subtracted and rearranged the order of its' zinc fingers, thereby allowing the creation of new populations of hotspots (Oliver et al. 2009; Baudat et al. 2010; Berg et al. 2010; Myers et al. 2010; Berg et al. 2011; Brick et al. 2012).

Organisms possessing *Prdm9* should see a rapid turnover of hotspots, and indeed, differences in DNA-binding appear to have created fine scale recombination rate divergence between individuals, populations, and species (Ptak et al. 2004; Ptak et al. 2005; Winckler et al. 2005; Jeffreys et al. 2009; Berg et al. 2010; Myers et al. 2010; Berg et al. 2011; Hinch et al. 2011; Smagulova et al. 2011; Auton et al. 2012; Brick et al. 2012). However, many organisms, including dogs, reptiles, birds, plants, yeast, nematodes, and Drosophila do not possess a copy of *Prdm9* (Birtle et al. 2006; Oliver et al. 2009; Ponting 2011; Heil et al. 2012). How are sites of recombination selected in these species? Does fine scale recombination rate turnover persist in the absence of *Prdm9*? Do fine scale recombination rates diverge even in species lacking hotspots?

Drosophila, for example, have an atypical recombination landscape in that they seem to lack hotspots, the only organism like this besides *C. elegans* (Hey 2004; Singh et al. 2009; Miller et al. 2012; Manzano-Winkler et al. 2013; Singh et al. 2013). Although they are missing this extreme fine scale variation in recombination rate, recombination rate is

known to vary over different scales and between populations and species (Brooks et al. 1986; True et al. 1996; Cirulli et al. 2007; Kulathinal et al. 2008; Stevison et al. 2010; Chan et al. 2012; Comeron et al. 2012; McGaugh et al. 2012). The unique recombination landscape, combined with a hundred years of recombination research, makes Drosophila an intriguing target to further explore.

In this study, we infer historic recombination rates from linkage disequilibrium present in modern-day DNA sequences using the procedure implemented in the program LDhelmet (Chan et al. 2012) to estimate fine scale recombination rates in *Drosophila pseudoobscura* and its closely related relative *Drosophila miranda* (~2-3 my divergence). We leverage empirical maps in both these species to validate our computational approach. We identify patterns of where recombination is happening in relation to genomic features. Finally, we compare recombination rates between *D. pseudoobscura* and *D. miranda* at multiple scales and deduce that recombination rates are more divergent at fine scales and conserved at broad scales, as seen in human-chimpanzee.

Materials and Methods

Population sequencing and variant calling

We used 11 whole genome sequences of *D. pseudoobscura* (Lines: MV2-25 (reference), Mather32, MSH24, MSH9, TL, PP1134, PP1137, AFC12, FLG14, FLG16,

FLG18) from McGaugh et al. (McGaugh et al. 2012) and 23 whole genome sequences of *D. miranda* (Lines: 0101.5, MA28, MA32, MAO101.4, MAO101.7, MAO3.1, MAO3.3, MAO3.4, MAO3.5, MAO3.6, ML1, ML14, ML16, ML22, ML26, ML4, ML51, ML6f, ML71, MSH22, MSH38, SP138, and SP235), provided by Doris Bachtrog's lab. All sequences were previously aligned to the *D. pseudoobscura* reference sequence v2.9.

The following steps were taken to call variants for the *D. miranda* genome sequences. (1) We used the Genome Analysis Tool Kit (GATK) v2.7-2 "RealignerTargetCreator" and "IndelRealigner" tools to locally realign reads and minimize mismatch calls due to the presence of insertions or deletions (McKenna et al. 2010; DePristo et al. 2011). (2) samtools v0.1.18 and bcftools v0.1.17 were used to call variants (Li et al. 2009). We filtered out unmapped reads and used a mapping quality score of 15. (3) We used a set of custom Perl scripts to convert the variant output files into individual chromosome aligned FASTA files for input to LDhelmet. It is imperative that each FASTA sequence in the aligned file is the same length.

Recombination estimates

Empirical estimates of recombination for *D. pseudoobscura* and *D. miranda* were obtained from McGaugh et al. (McGaugh et al. 2012). Population sequencing based estimates of recombination were determined using the LDhelmet program (Chan et al. 2012). LDhelmet is modeled after LDhat (McVean et al. 2002; McVean et al. 2004), but

specifically tailored to address issues in Drosophila, such as a magnitude fold higher background recombination rate, higher SNP density, and a large portion of the genome influenced by positive selection (Sella et al. 2009; Sattath et al. 2011). The program was run individually for each chromosome using default/suggested parameters, with the exception that we estimated theta from our dataset and created our own mutation transition matrix. A block penalty of 50 was used unless otherwise noted.

To compare empirical recombination estimates to LD-based recombination estimates the recombination estimate from LDhelmet was corrected for distance and averaged over a given interval. The empirical chromosomal recombination average (cM/Mb) was divided by the total average LD-based recombination rate for a chromosome to get a conversion factor. Each interval's average LD-based estimate was multiplied by the conversion factor to get an approximation of recombination rates in the units centiMorgan per Megabase (cM/Mb) per Chan *et. al* (2012). Because we were interested in comparing recombination rates between species, and empirical data suggests that *D. miranda* has a higher recombination rate than *D. pseudoobscura* (McGaugh et al. 2012), we sought an additional method of converting LD-based recombination rates without assuming knowledge of empirical recombination rates.

Analyses of genomic correlates

Transcription start sites (TSS): Locations of *D. pseudoobscura* transcription start sites (TSS) were obtained from Main *et al.* (Main et al. 2013). The recombination estimate from LDhelmet was corrected for distance and averaged over 5000bp intervals for the 200kb regions flanking a TSS. This data was aggregated together to create chromosome specific TSS figures. **Introns and Exons**: Locations of exons and introns and relative positions in a gene were extracted from *D. pseudoobscura* v2.9 annotations from FlyBase (St Pierre et al. 2014). The recombination estimate from LDhelmet was corrected for distance and averaged over the given interval, then aggregated to give chromosome and genome wide totals for each exon and intron position within a gene. **Nucleotide diversity**: For our analyses of nucleotide diversity, FlyBase annotations of the *D. pseudoobscura* genome v.2.9 and Perl scripts were used to identify four-fold degenerate coding positions. We created Perl scripts to calculate pairwise nucleotide diversity (π) using four-fold degenerate sites by making all pairwise comparisons for each given interval (e.g. for D. pseudoobscura diversity all pairwise comparisons between the 11 lines were made and for D. miranda, all pairwise comparisons between the 23 lines were made), excluding sites where an insertion or deletion was found in any line. The number of SNP bases was averaged across each recombination interval for all pairwise comparisons.

Results and Discussion

LD-based recombination maps are good estimates of empirical recombination maps

Historically researchers have used a number of methods to estimate recombination rate including sperm-typing (Jeffreys et al. 1998; Jeffreys et al. 2001), genotyping pedigrees or controlled crosses (Kong et al. 2010), using immunocytological markers (Dumont et al. 2011), ChIP-seq of proteins associated with double strand break or crossover intermediates (Pan et al. 2011; Brick et al. 2012), and employing statistical methods to population sequencing data (Myers et al. 2005). It has been particularly difficult to achieve fine scale resolution in organisms besides human, mouse, and yeast. The latter approach was believed to be inappropriate in Drosophila, where background recombination rates are much higher than humans, for which it was created. A slew of other problems including higher SNP density and known genome wide effects of positive selection (Sella et al. 2009; Sattath et al. 2011) added to the complexity.

Therefore, to create fine scale recombination maps in *D. pseudoobscura* and *D. miranda*, we utilized a new program: LDhelmet (Chan et al. 2012), a statistical approach designed for Drosophila that estimates a population recombination rate, ρ, from population sequencing data as in (McVean et al. 2002; McVean et al. 2004). The program calculates $\rho = 4N_e r$ (where N_e is the effective population size and r is the recombination rate per generation) to estimate the amount of recombination needed in the population to produce the observed levels of linkage disequilibrium under a given model. Linkage

disequilibrium can be influenced by processes besides recombination including mutation, genetic drift, selection, demography, migration, and gene conversion (Slatkin 2008), and some studies have identified discrepancies between empirical and LD-based recombination maps (Jeffreys et al. 2005; Kauppi et al. 2005; Reed et al. 2006).

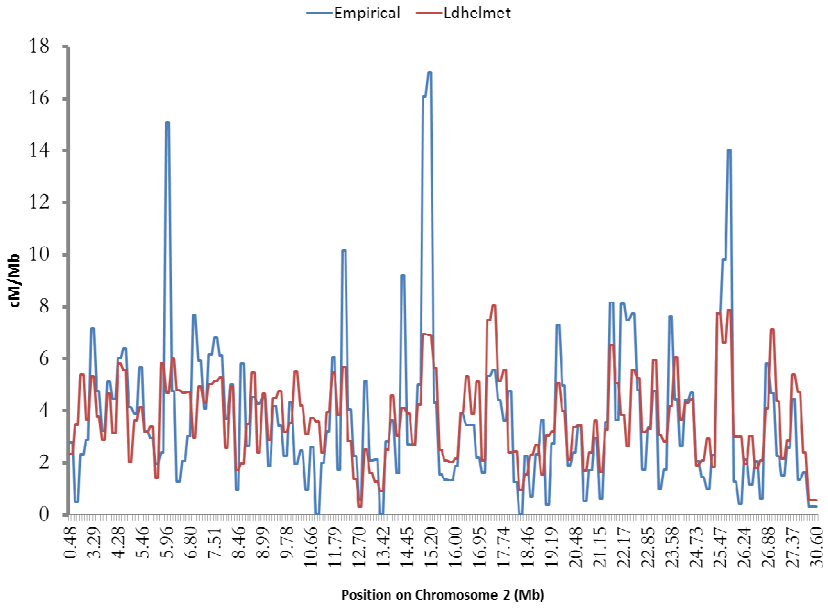

Figure 12: Comparison of *D. pseudoobscura* empirical and LDhelmet recombination rates on chromosome 2

To identify how well our LD-based maps replicates empirical recombination rates, we used empirical recombination rates estimated from two populations of *D. pseudoobscura* and one population of *D. miranda* (McGaugh et al. 2012). Empirical and LD-based recombination estimates are positively and significantly correlated at broad scales (**Figure 12**; ~150kb intervals; Chr 2: $p<0.0001$, $r=0.6266$; X chr: $p<0.0001$, $r=0.7114$) and fine scales (~20 kb intervals, $p=0.0084$, $r=0.6336$) across the genome in *D. pseudoobscura* . The broad-scale correlation is similar to that of the two empirical *D.*

pseudoobscura population maps (r=0.7183), suggesting that we are capturing the full amount of species variation in our LD-based map. In *D. miranda*, the effect is less pronounced, although still significant (X chr: p=0.0090, r=0.3989). It is unclear why the correlation between the empirical and LD-based estimates is weaker in *D. miranda*. One possibility is that there is greater population variation in *D. miranda*, as seen in *D. melanogaster* (Comeron et al. 2012), although without another empirical map to compare, it's not possible to make conclusions.

Between species recombination rate comparison

Previous work has shown that recombination rates are conserved at broad scales between closely related species of mammals, birds, and insects (Smukowski et al. 2011). In contrast, it's clear that fine scale recombination rates diverge rapidly in humans (Berg et al. 2010; Myers et al. 2010; Berg et al. 2011; Hinch et al. 2011; Auton et al. 2012), but little data exists outside primates and it remains uncertain if this phenomenon applies across taxa. In yeast, hotspot turnover is indeed less rapid. Hotspots show relative conservation (about 50%) between *Saccharomyces cerevisae* and *Saccharomyces paradoxus* despite 13% sequence divergence, although this is likely attributable to the low frequency of sex and outcrossing in these species (Tsai et al. 2010). In dogs, which have a non-functional copy of *Prdm9*, conclusions are mixed. Inferring recombination hotspot history from GC-biased gene conversion, researchers have come to differing opinions

regarding the evolutionary stability of hotspots in this species (Axelsson et al. 2012; Auton et al. 2013). Regardless, no true comparison has been made in dogs to closely related species, so this remains an intriguing question.

Recombination rate is known to be conserved at broad scales (~200kb) between *D. pseudoobscura* and *D. miranda*, although rates in *D. miranda* are significantly elevated across all regions surveyed (McGaugh et al. 2012). With our LD-based recombination maps, we have the opportunity to identify if the putative global recombination modifier identified in empirical work persists species-wide in *D. miranda*. Indeed, we recapitulate results of McGaugh *et al*., detecting broad scale conservation of recombination rates between species, with *D. miranda* recombination rates remaining elevated (X chr: 1 Mb: $r=0.7586$; 100kb: $p<0.0001$, $r=0.6391$). Moreover, *D. miranda* recombination rates are higher than detected in empirical work, with average rho values of approximately three times that of *D. pseudoobscura* on the X chromosome. This difference persists despite a difference in effective population sizes, *D. pseudoobscura* N_e is thought to be several times that of *D. miranda* (Bachtrog et al. 2006; Loewe et al. 2006; Jensen et al. 2011). We can only speculate as to why recombination rates would increase in *D. miranda* (or decrease in D. *pseudoobscura*) since their divergence. Theoretical work shows selection for or against new haplotypes can indirectly select for alleles that modify recombination rates (Otto et al. 2002), observations in domesticated crops and several direct selection experiments in

Drosophila support this (Flexon et al. 1982; Burt et al. 1987; Korol et al. 1994; Otto et al. 2001; Ross-Ibarra 2004; Coop et al. 2007).

Analysis of fine scale recombination in Drosophila has been limited to isolated genomic regions in *D. melanogaster* (Schweitzer 1935; Singh et al. 2009; Miller et al. 2012; Singh et al. 2013) and *D. pseudoobscura* (Cirulli et al. 2007; McGaugh et al. 2012; Manzano-Winkler et al. 2013). With LD-based maps, we are able to characterize the genome wide fine scale recombination landscape in *D. pseudoobscura* and *D. miranda* for the first time, and to test whether turnover in fine scale recombination rates occurs outside of mammals. Recombination rates averaged over 10kb windows show heterogeneity in both species, and again, *D. miranda* has higher recombination than *D. pseudoobscura*. However, we detect less variation in *D. miranda* than *D. pseudoobscura* (**Figure 13**). Whether this is an artifact of lower nucleotide diversity in *D. miranda* is unclear. We do detect greater divergence at fine scales (X chr: 10kb: $p<0.0001$, $r=0.4604$), consistent with less constraint on recombination rates or greater error in estimation at this scale. It's suggestive that Drosophila too face the hotspot paradox, and as they don't have a *Prdm9*-like protein, have come up with an alternate solution to counteract the loss.

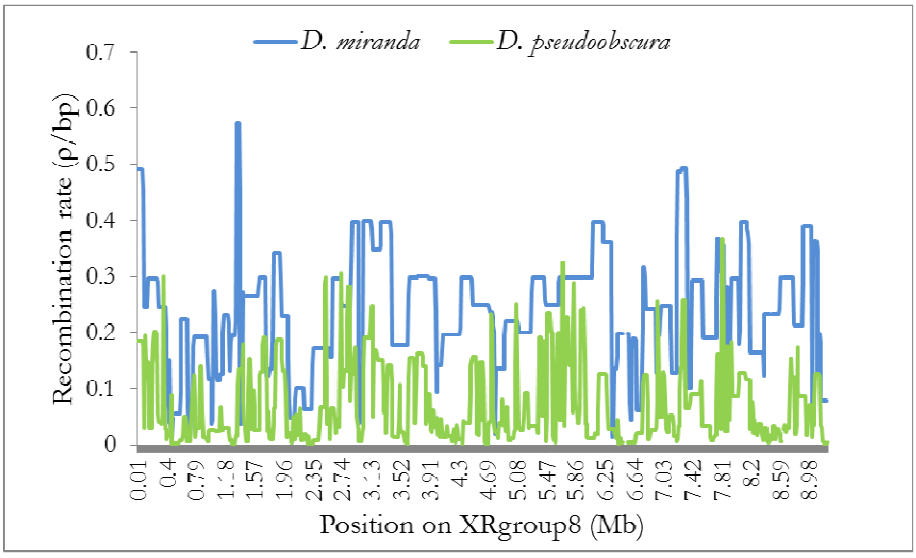

Figure 13: Fine scale recombination rates (10kb) between species

Distribution of recombination rate across the genome

Transcription Start Sites and coding regions

Hotspots in yeast, plants, and dog are primarily found overlapping promoters, presumably due to the open chromatin structure at promoters allowing recombination machinery access to the DNA (Petes 2001; Lichten 2008; Mancera et al. 2008; Pan et al. 2011; Choi et al. 2013; Hellsten et al. 2013). In contrast, hotspots primarily occur outside of promoters in mouse and human (Myers et al. 2005; Smagulova et al. 2011), and there is evidence that *Prdm9* may even be directing recombination machinery away from promoters and other functional genomic elements (Brick et al. 2012). In these species, interactions with histones, particularly a meiosis specific H3K4me3 modification , govern hotspot distribution. These observations set up a model in which organisms

without *Prdm9* (yeast, plants, dog) are opportunistic: they form recombination hotspots in "windows of opportunity," most often nucleosome depleted regions around promoters. On the other hand, in organisms with *Prdm9* (humans, mouse), *Prdm9* most likely functions to introduce meiotic specific H3K4me3 modifications and directs recombination machinery to initiate at these locations away from genomic elements (Hayashi et al. 2005; Grey et al. 2011). Although factors that determine hotspots in these organisms differ, the overall patterns and intensities of hotspots are remarkably similar between organisms.

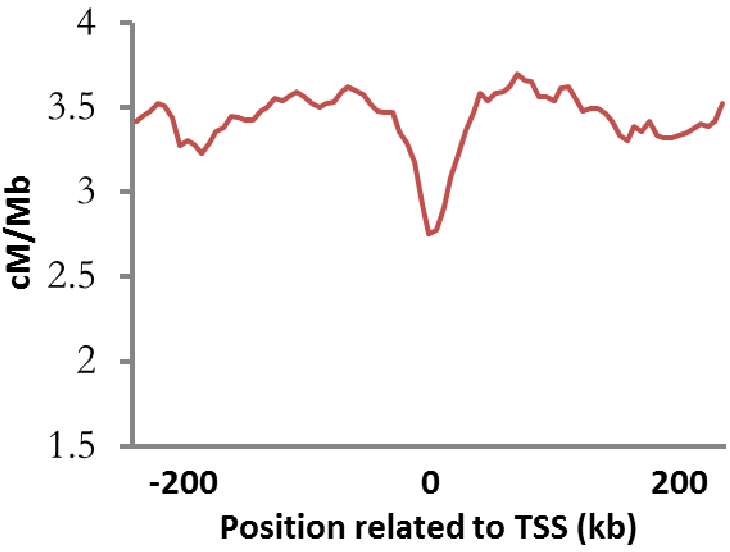

Figure 14: Recombination rate around transcription start sites

In *D. pseudoobscura*, recombination is reduced around transcription start sites and adjacent regions up to 35 kb away (**Figure 14**). This is supported by findings in *D.*

melanogaster as well (Chan et al. 2012), although the reduction in rates appears to persist for greater distances in *D. pseudoobscura*. Similarly, we see reduction in recombination rates at the 5' end of genes, with greater rates in introns than exons, which both increase with distance from the start of the coding region (**Figure 15**). This reveals a contradictory pattern from the monkeyflower, *Mimulus guttatus* (Hellsten et al. 2013), and speaks to Drosophila's unusual recombination landscape. It remains unclear what genomic features influence the distribution of recombination events in Drosophila; future studies looking at histone modifications and nucleosome occupancy will hopefully shed light on this enigma.

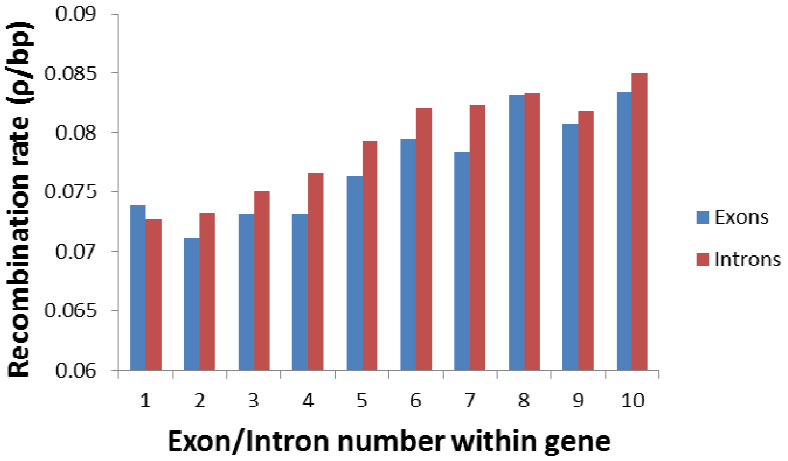

Figure 15: Recombination rate within genes

Cumulative recombination map

Looking at the proportion of recombination that happens in a proportion of the sequence reveals the degree to which recombination events cluster across the genome. In humans, recombination is highly punctate, with about 80% of recombination happening in less than 20% of the sequence (Myers et al. 2006), or a Gini coefficient of approximately 0.8 (Kong et al. 2010; Kaur et al. 2014). As previously discussed, Drosophila lack hotspots characteristic of other organisms (Singh et al. 2009; Manzano-Winkler et al. 2013; Singh et al. 2013), although fine scale variation has been identified (see above), and Chan *et al.* even detected a handful of more typical hotspots (Chan et al. 2012). This is reflected in *D. pseudoobscura*'s intermediate Gini coefficient of approximately 0.50, or about 80% of recombination occurring in 50% of the sequence (**Figure 16**). This is similar to the Gini coeffiecient estimated for *D. melanoagaster* (0.47), falling between *C. elegans* (0.28) and *S. cerevisae* (0.64) (Kaur et al. 2014).

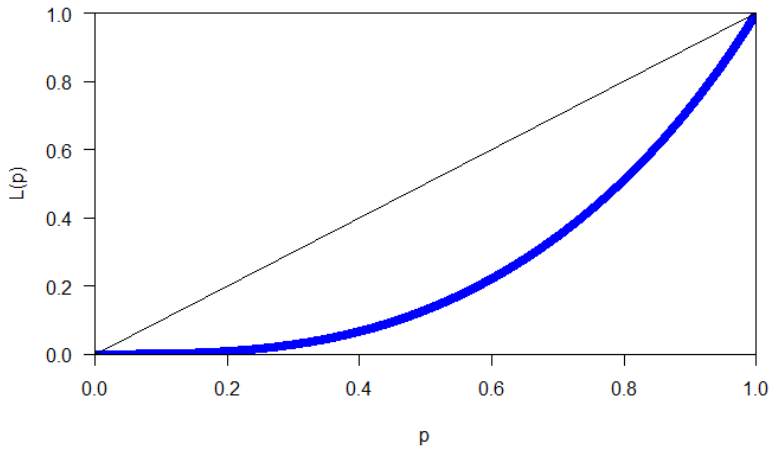

Figure 16: Proportion recombination occurring in proportion of sequence

Conclusions

Here we employ a computational approach to present the first genome wide fine scale recombination maps for *D. pseudoobscura* and *D. miranda*. Utilizing existing empirical recombination maps in these species, we were able to validate our results and conclude the program LDhelmet is a valuable resource for the Drosophila and recombination communities. Furthermore, we reveal an unprecedented view of the distribution of recombination in these species, illustrating new patterns of where recombination is taking place. Finally, we show broad scale conservation of recombination rates and fine scale divergence between closely related species, one of the only studies to show this phenomenon exists outside of mouse and human. This raises interesting implications about the evolution of meiosis and selection pressures on recombination that persist across distantly related taxa with widely different recombination landscapes.

Future studies can apply these types of fine scale recombination maps to identify determinants of recombination, improve theory and modeling, and to better understand molecular evolutionary processes and genome evolution.

Works Cited

Acquaviva, L., L. Szekvolgyi, et al. (2013). "The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination." Science **339**(6116): 215-218.

Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of Drosophila melanogaster." Science **287**(5461): 2185-2195.

Aguade, M., N. Miyashita, et al. (1989). "Reduced variation in the yellow-achaete-scute region in natural populations of Drosophila melanogaster." Genetics **122**(3): 607-615.

Altschul, S. F., W. Gish, et al. (1990). "Basic Local Alignment Search Tool." Journal of Molecular Biology **215**(3): 403-410.

Altshuler, D. L., R. M. Durbin, et al. (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

Andolfatto, P. (2007). "Hitchhiking effects of recurrent beneficial amino acid substitutions in the Drosophila melanogaster genome." Genome Res **17**(12): 1755-1762.

Aquadro, C. F., V. B. DuMont, et al. (2001). "Genome-wide variation in the human and fruitfly: a comparison." Current Opinion in Genetics & Development **11**(6): 627-634.

Aravin, A. A., G. J. Hannon, et al. (2007). "The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race." Science **318**(5851): 761-764.

Ardlie, K. G., L. Kruglyak, et al. (2002). "Patterns of linkage disequilibrium in the human genome." Nature Reviews Genetics **3**(4): 299-309.

Arguello, J. R., Y. Zhang, et al. (2010). "Recombination Yet Inefficient Selection along the Drosophila melanogaster Subgroup's Fourth Chromosome." Molecular Biology and Evolution **27**(4): 848-861.

Auton, A., A. Fledel-Alon, et al. (2012). "A Fine-Scale Chimpanzee Genetic Map from Population Sequencing." Science **336**(6078): 193-198.

Auton, A., Y. Rui Li, et al. (2013). "Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs." PLoS Genet **9**(12): e1003984.

Axelsson, E., M. T. Webster, et al. (2012). "Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome." Genome Research **22**(1): 51-63.

Bachtrog, D. (2003). "Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in Drosophila." Nat Genet **34**(2): 215-219.

Bachtrog, D. and P. Andolfatto (2006). "Selection, recombination and demographic history in Drosophila miranda." Genetics **174**(4): 2045-2059.

Bachtrog, D. and B. Charlesworth (2002). "Reduced adaptation of a non-recombining neo-Y chromosome." Nature **416**(6878): 323-326.

Backstrom, N., W. Forstmeier, et al. (2010). "The recombination landscape of the zebra finch Taeniopygia guttata genome." Genome Research **20**(4): 485-495.

Ballestar, E., L. A. Pile, et al. (2001). "A Drosophila MBD family member is a transcriptional corepressor associated with specific genes." Eur J Biochem **268**(20): 5397-5406.

Barthes, P., J. Buard, B. de Massy (2011). Epigenetic Factors and Regulation of Meiotic Recombination in Mammals Springer.

Bartolome, C. and B. Charlesworth (2006). "Rates and patterns of chromosomal evolution in Drosophila pseudoobscura and D. miranda." Genetics **173**(2): 779-791.

Bartolome, C., X. Maside, et al. (2005). "Patterns of selection on synonymous and nonsynonymous variants in Drosophila miranda." Genetics **169**(3): 1495-1507.

Barton, N. H. (1995). "A General-Model for the Evolution of Recombination." Genetical Research **65**(2): 123-144.

Barton, N. H. (2010). "Mutation and the evolution of recombination." Philosophical Transactions of the Royal Society B-Biological Sciences **365**(1544): 1281-1294.

Baudat, F., J. Buard, et al. (2010). "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice." Science **327**(5967): 836-840.

Baudry, E., C. Kerdelhue, et al. (2001). "Species and recombination effects on DNA variability in the tomato genus." Genetics **158**(4): 1725-1735.

Beavis, W. D. (1996). "QTL mapping in plant breeding populations." American Statistical Association 1996 Proceedings of the Biometrics Section: 13-20
431.

Beavis, W. D., O. S. Smith, et al. (1994). "Identification of Quantitative Trait Loci Using a Small Sample of Topcrossed and F4 Progeny from Maize." Crop Science **34**(4): 882-896.

Becker, M., S. Muller, et al. (2012). "Pmt1, a Dnmt2 homolog in Schizosaccharomyces pombe, mediates tRNA methylation in response to nutrient signaling." Nucleic Acids Research **40**(22): 11648-11658.

Begun, D. J. and C. F. Aquadro (1992). "Levels of Naturally-Occurring DNA Polymorphism Correlate with Recombination Rates in Drosophila-Melanogaster." Nature **356**(6369): 519-520.

Begun, D. J., A. K. Holloway, et al. (2007). "Population genomics: Whole-genome analysis of polymorphism and divergence in Drosophila simulans." Plos Biology **5**(11): 2534-2559.

Benjamini, Y., and Y Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." Journal of the Royal Statistical Society: Series B **57**: 289-300.

Bennetzen, J. L., K. Schrick, et al. (1994). "Active Maize Genes Are Unmodified and Flanked by Diverse Classes of Modified, Highly Repetitive DNA." Genome **37**(4): 565-576.

Berchowitz, L. E., S. E. Hanlon, et al. (2009). "A positive but complex association between meiotic double-strand break hotspots and open chromatin in Saccharomyces cerevisiae." Genome Res **19**(12): 2245-2257.

Berchowitz, L. E., S. E. Hanlon, et al. (2009). "A positive but complex association between meiotic double-strand break hotspots and open chromatin in Saccharomyces cerevisiae." Genome Research **19**(12): 2245-2257.

Berg, I. L., R. Neumann, et al. (2010). "PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans." Nature Genetics **42**(10): 859-+.

Berg, I. L., R. Neumann, et al. (2011). "Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations." Proceedings of the National Academy of Sciences of the United States of America **108**(30): 12378-12383.

Betancourt, A. J. and D. C. Presgraves (2002). "Linkage limits the power of natural selection in Drosophila." Proc Natl Acad Sci U S A **99**(21): 13616-13620.

Betancourt, A. J., J. J. Welch, et al. (2009). "Reduced effectiveness of selection caused by a lack of recombination." Curr Biol **19**(8): 655-660.

Beukeboom, L. W., O. Niehuis, et al. (2010). "A comparison of recombination frequencies in intraspecific versus interspecific mapping populations of Nasonia." Heredity **104**(3): 302-309.

Beye, M., I. Gattermeier, et al. (2006). "Exceptionally high levels of recombination across the honey bee genome." Genome Research **16**(11): 1339-1344.

Birky, C. W., Jr. and J. B. Walsh (1988). "Effects of linkage on rates of molecular evolution." Proc Natl Acad Sci U S A **85**(17): 6414-6418.

Birtle, Z. and C. P. Ponting (2006). "Meisetz and the birth of the KRAB motif." Bioinformatics **22**(23): 2841-2845.

Blumenstiel, J. P. (2011). "Evolutionary dynamics of transposable elements in a small RNA world." Trends Genet **27**(1): 23-31.

Borde, V., N. Robine, et al. (2009). "Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites." Embo Journal **28**(2): 99-111.

Boulton, A., R. S. Myers, et al. (1997). "The hotspot conversion paradox and the evolution of meiotic recombination." Proceedings of the National Academy of Sciences of the United States of America **94**(15): 8058-8063.

Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." Nature **431**(7004): 96-99.

Bourc'his, D., G. L. Xu, et al. (2001). "Dnmt3L and the establishment of maternal genomic imprints." Science **294**(5551): 2536-2539.

Bourguet, D., J. Gair, et al. (2003). "Genetic recombination and adaptation to fluctuating environments: selection for geotaxis in Drosophila melanogaster." Heredity (Edinb) **91**(1): 78-84.

Brennecke, J., A. A. Aravin, et al. (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila." Cell **128**(6): 1089-1103.

Brick, K., F. Smagulova, et al. (2012). "Genetic recombination is directed away from functional genomic elements in mice." Nature **485**(7400): 642-645.

Brooks, L. D. and R. W. Marks (1986). "The Organization of Genetic-Variation for Recombination in Drosophila-Melanogaster." Genetics **114**(2): 525-547.

Brown, T. C. and J. Jiricny (1987). "A Specific Mismatch Repair Event Protects Mammalian-Cells from Loss of 5-Methylcytosine." Cell **50**(6): 945-950.

Brown, T. C., I. Zbinden, et al. (1989). "Modified Sv40 for Analysis of Mismatch Repair in Simian and Human-Cells." Mutation Research **220**(2-3): 115-123.

Buard, J., P. Barthes, et al. (2009). "Distinct histone modifications define initiation and repair of meiotic recombination in the mouse." EMBO J **28**(17): 2616-2624.

Buhler, C., V. Borde, et al. (2007). "Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in Saccharomyces cerevisiae." Plos Biology **5**(12): 2797-2808.

Bullaughey, K., M. Przeworski, et al. (2008). "No effect of recombination on the efficacy of natural selection in primates." Genome Res **18**(4): 544-554.

Burt, A. and G. Bell (1987). "Mammalian Chiasma Frequencies as a Test of 2 Theories of Recombination." Nature **326**(6115): 803-805.

Bussell, J. J., N. M. Pearson, et al. (2006). "Human polymorphism and human-chimpanzee divergence in pseudoautosomal region correlate with local recombination rate." Gene **368**: 94-100.

Campos, J. L., B. Charlesworth, et al. (2012). "Molecular evolution in nonrecombining regions of the Drosophila melanogaster genome." Genome Biology and Evolution **4**(3): 278-288.

Carrington, M. and M. Cullen (2004). "Justified chauvinism: advances in defining meiotic recombination through sperm typing." Trends in Genetics **20**(4): 196-205.

Cedar, H. and Y. Bergman (2009). "Linking DNA methylation and histone modification: patterns and paradigms." Nat Rev Genet **10**(5): 295-304.

Chan, A. H., P. A. Jenkins, et al. (2012). "Genome-wide fine-scale recombination rate variation in Drosophila melanogaster." PLoS Genet **8**(12): e1003090.

Charlesworth, B. (1990). "Mutation-Selection Balance and the Evolutionary Advantage of Sex and Recombination." Genetical Research **55**(3): 199-221.

Charlesworth, B. (2010). "Evolution. Variation catches a ride." Science **330**(6002): 326-327.

Charlesworth, B. (2012). "The effects of deleterious mutations on evolution at linked sites." Genetics **190**(1): 5-22.

Charlesworth, B., A. J. Betancourt, et al. (2009). "Genetic recombination and molecular evolution." Cold Spring Harb Symp Quant Biol **74**: 177-186.

Charlesworth, B. and D. Charlesworth (1985). "Genetic-Variation in Recombination in Drosophila .1. Responses to Selection and Preliminary Genetic-Analysis." Heredity **54**(Feb): 71-83.

Charlesworth, B. and D. Charlesworth (2000). "The degeneration of Y chromosomes." Philos Trans R Soc Lond B Biol Sci **355**(1403): 1563-1572.

Charlesworth, B., M. T. Morgan, et al. (1993). "The Effect of Deleterious Mutations on Neutral Molecular Variation." Genetics **134**(4): 1289-1303.

Charlesworth, B., P. Sniegowski, et al. (1994). "The Evolutionary Dynamics of Repetitive DNA in Eukaryotes." Nature **371**(6494): 215-220.

Cheng, E. Y., P. A. Hunt, et al. (2009). "Meiotic Recombination in Human Oocytes." Plos Genetics **5**(9): -.

Chinnici, J. P. (1971). "Modification of Recombination Frequency in Drosophila .1. Selection for Increased and Decreased Crossing Over." Genetics **69**(1): 71-&.

Chinnici, J. P. (1971). "Modification of Recombination Frequency in Drosophila .2. Polygenic Control of Crossing Over." Genetics **69**(1): 85-&.

Chodavarapu, R. K., S. Feng, et al. (2010). "Relationship between nucleosome positioning and DNA methylation." Nature **466**(7304): 388-392.

Choi, K., X. Zhao, et al. (2013). "Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters." Nat Genet **45**(11): 1327-1336.

Cirulli, E. T., R. M. Kliman, et al. (2007). "Fine-scale crossover rate heterogeneity in Drosophila pseudoobscura." Journal of Molecular Evolution **64**(1): 129-135.

Clark, A. G., M. B. Eisen, et al. (2007). "Evolution of genes and genomes on the Drosophila phylogeny." Nature **450**(7167): 203-218.

Clark, S. H., A. J. Hilliker, et al. (1988). "Recombination can initiate and terminate at a large number of sites within the rosy locus of Drosophila melanogaster." Genetics **118**(2): 261-266.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences Hillsdale,NJ, Lawrence Erlbaum.

Colegrave, N. (2002). "Sex releases the speed limit on evolution." Nature **420**(6916): 664-666.

Comeron, J. M. and M. Kreitman (2002). "Population, evolutionary and genomic consequences of interference selection." Genetics **161**(1): 389-410.

Comeron, J. M., M. Kreitman, et al. (1999). "Natural selection on synonymous sites is correlated with gene length and recombination in Drosophila." Genetics **151**(1): 239-249.

Comeron, J. M., R. Ratnappan, et al. (2012). "The Many Landscapes of Recombination in Drosophila melanogaster." Plos Genetics **8**(10).

Coop, G. and M. Przeworski (2007). "An evolutionary view of human recombination." Nature Reviews Genetics **8**(1): 23-34.

Coop, G., X. Q. Wen, et al. (2008). "High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans." Science **319**(5868): 1395-1398.

Copenhaver, G. P. (2005). "Plant genetics: When not to interfere." Current Biology **15**(8): R290-R291.

Crooks, G. E., G. Hon, et al. (2004). "WebLogo: A sequence logo generator." Genome Research **14**(6): 1188-1190.

Cutter, A. D. and J. Y. Choi (2010). "Natural selection shapes nucleotide polymorphism across the genome of the nematode Caenorhabditis briggsae." Genome Res **20**(8): 1103-1111.

Cutter, A. D. and A. M. Moses (2011). "Polymorphism, divergence, and the role of recombination in Saccharomyces cerevisiae genome evolution." Molecular Biology and Evolution **28**(5): 1745-1754.

Cutter, A. D. and B. A. Payseur (2003). "Selection at linked sites in the partial selfer Caenorhabditis elegans." Molecular Biology and Evolution **20**(5): 665-673.

D'Avila, M. F., R. N. Garcia, et al. (2010). "Sex-specific methylation in Drosophila: an investigation of the Sophophora subgenus." Genetica **138**(8): 907-913.

Dawson, D. A., M. Akesson, et al. (2007). "Gene order and recombination rate in homologous chromosome regions of the chicken and a passerine bird." Molecular Biology and Evolution **24**(7): 1537-1552.

DePristo, M. A., E. Banks, et al. (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." Nat Genet **43**(5): 491-498.

Dobzhansky, T. (1973). "Is there gene exchange between Drosophila pseudoobsura and Drosophila persimilis in their natural habitats?" American Naturalist **107**: 312-314.

Dong, A., J. A. Yoder, et al. (2001). "Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA." Nucleic Acids Research **29**(2): 439-448.

Dumont, B. L., K. W. Broman, et al. (2009). "Variation in Genomic Recombination Rates Among Heterogeneous Stock Mice." Genetics **182**(4): 1345-1349.

Dumont, B. L. and B. A. Payseur (2008). "Evolution of the genomic rate of recombination in mammals." Evolution **62**(2): 276-294.

Dumont, B. L. and B. A. Payseur (2011). "Evolution of the genomic recombination rate in murid rodents." Genetics **187**(3): 643-657.

Dumont, B. L., M. A. White, et al. (2011). "Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps." Genome Research **21**(1): 114-125.

Durdevic, Z., K. Hanna, et al. (2013). "Efficient RNA virus control in Drosophila requires the RNA methyltransferase Dnmt2." EMBO Rep **14**(3): 269-275.

Durdevic, Z., M. B. Mobin, et al. (2013). "The RNA methyltransferase Dnmt2 is required for efficient Dicer-2-dependent siRNA pathway activity in Drosophila." Cell Rep **4**(5): 931-937.

Durdevic, Z. and M. Schaefer (2013). "Dnmt2 methyltransferases and immunity: an ancient overlooked connection between nucleotide modification and host defense?" Bioessays **35**(12): 1044-1049.

Duret, L. and P. F. Arndt (2008). "The impact of recombination on nucleotide substitutions in the human genome." Plos Genetics **4**(5): -.

Ellermeier, C., E. C. Higuchi, et al. (2010). "RNAi and heterochromatin repress centromeric meiotic recombination." Proceedings of the National Academy of Sciences of the United States of America **107**(19): 8701-8705.

Esposito, M. S. and C. V. Bruschi (1993). "Diploid Yeast-Cells Yield Homozygous Spontaneous Mutations." Current Genetics **23**(5-6): 430-434.

Fan, J. B., A. Oliphant, et al. (2003). "Highly parallel SNP genotyping." Cold Spring Harb Symp Quant Biol **68**: 69-78.

Fan, J. B., A. Oliphant, et al. (2003). Highly parallel SNP genotyping.

Feldman, M. W., S. P. Otto, et al. (1996). "Population genetic perspectives on the evolution of recombination." Annual Review of Genetics **30**: 261-295.

Feng, S., S. J. Cokus, et al. (2010). "Conservation and divergence of methylation patterning in plants and animals." Proc Natl Acad Sci U S A **107**(19): 8689-8694.

Filatov, D. A. and D. T. Gerrard (2003). "High mutation rates in human and ape pseudoautosomal genes." Gene **317**(1-2): 67-77.

Fitzpatrick, C. L., L. Stevison, M. Noor (2009). "Fine-scale crossover rate and interference along the XR-chromosome arm of Drosophila pseudoobscura." Drosophila Information Service **92**: 27–29.

Fledel-Alon, A., D. J. Wilson, et al. (2009). "Broad-Scale Recombination Patterns Underlying Proper Disjunction in Humans." Plos Genetics **5**(9): -.

Flexon, P. B. and C. F. Rodell (1982). "Genetic recombination and directional selection for DDT resistance in Drosophila melanogaster." Nature **298**(5875): 672-674.

Fog, C. K., G. G. Galli, et al. (2012). "PRDM proteins: Important players in differentiation and disease." Bioessays **34**(1): 50-60.

Foss, E., R. Lande, et al. (1993). "Chiasma Interference as a Function of Genetic-Distance." Genetics **133**(3): 681-691.

Fu, F., J. D. Sander, et al. (2009). "Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays." Nucleic Acids Research **37**: D279-D283.

Fumasoni, I., N. Meani, et al. (2007). "Family expansion and gene rearrangements contributed to the functional specialization of PRDM genes in vertebrates." Bmc Evolutionary Biology **7**.

Gerton, J. L., J. DeRisi, et al. (2000). "Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae." Proceedings of the National Academy of Sciences of the United States of America **97**(21): 11383-11390.

Gloor, G. B. a. W. R. E. (1992). "Single-fly DNA preps for PCR." Drosophila Information Service **71**: 148-149.

Goddard, M. R., H. C. Godfray, et al. (2005). "Sex increases the efficacy of natural selection in experimental yeast populations." Nature **434**(7033): 636-640.

Gowher, H., O. Leismann, et al. (2000). "DNA of Drosophila melanogaster contains 5-methylcytosine." EMBO J **19**(24): 6918-6923.

Graffelman, J., D. J. Balding, et al. (2007). "Variation in estimated recombination rates across human populations." Human Genetics **122**(3-4): 301-310.

Grey, C., P. Barthes, et al. (2011). "Mouse PRDM9 DNA-Binding Specificity Determines Sites of Histone H3 Lysine 4 Trimethylation for Initiation of Meiotic Recombination." Plos Biology **9**(10).

Groenen, M. A. M., P. Wahlberg, et al. (2009). "A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate." Genome Research **19**(3): 510-519.

Guillon, H. and B. de Massy (2002). "An initiation site for meiotic crossing-over and gene conversion in the mouse." Nature Genetics **32**(2): 296-299.

Haddrill, P. R., D. L. Halligan, et al. (2007). "Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over." Genome Biol **8**(2): R18.

Haddrill, P. R., L. Loewe, et al. (2010). "Estimating the parameters of selection on nonsynonymous mutations in Drosophila pseudoobscura and D. miranda." Genetics **185**(4): 1381-1396.

Haddrill, P. R., F. M. Waldron, et al. (2008). "Elevated levels of expression associated with regions of the Drosophila genome that lack crossing over." Biol Lett **4**(6): 758-761.

Hahn, M. W. (2008). "Toward a selection theory of molecular evolution." Evolution **62**(2): 255-265.

Halligan, D. L., F. Oliver, et al. (2010). "Evidence for pervasive adaptive protein evolution in wild mice." PLoS Genet **6**(1): e1000825.

Hassold, T. and P. Hunt (2001). "To err (meiotically) is human: the genesis of human aneuploidy." Nat Rev Genet **2**(4): 280-291.

Hassold, T., L. Judis, et al. (2004). "Cytological studies of meiotic recombination in human males." Cytogenetic and Genome Research **107**(3-4): 249-255.

Hassold, T., S. Sherman, et al. (2000). "Counting cross-overs: characterizing meiotic recombination in mammals." Human Molecular Genetics **9**(16): 2409-2419.

Hawley, R., D. T. Harris, et al. (2002). "Meiotic chromosome segregation in Drosophila." Molecular Biology of the Cell **13**: 150a-150a.

Hawley, R. S., H. Irick, et al. (1992). "There Are 2 Mechanisms of Achiasmate Segregation in Drosophila Females, One of Which Requires Heterochromatic Homology." Developmental Genetics **13**(6): 440-467.

Hayashi, K., K. Yoshida, et al. (2005). "A histone H3 methyltransferase controls epigenetic events required for meiotic prophase." Nature **438**(7066): 374-378.

Heil, C. S. and M. A. Noor (2012). "Zinc finger binding motifs do not explain recombination rate variation within or between species of Drosophila." Plos One **7**(9): e45055.

Hellmann, I., I. Ebersberger, et al. (2003). "A neutral explanation for the correlation of diversity with recombination rates in humans." American Journal of Human Genetics **72**(6): 1527-1535.

Hellsten, U., K. M. Wright, et al. (2013). "Fine-scale variation in meiotic recombination in Mimulus inferred from population shotgun sequencing." Proc Natl Acad Sci U S A **110**(48): 19478-19482.

Hey, J. (2004). "What's so hot about recombination hotspots?" Plos Biology **2**(6): 730-733.

Hey, J. and R. M. Kliman (2002). "Interactions between natural selection, recombination and gene density in the genes of drosophila." Genetics **160**(2): 595-608.

Hicks, W. M., M. Kim, et al. (2010). "Increased mutagenesis and unique mutation signature associated with mitotic gene conversion." Science **329**(5987): 82-85.

Hill, W. G. and Robertso.A (1966). "Effect of Linkage on Limits to Artificial Selection." Genetical Research **8**(3): 269-&.

Hillers, K. J. (2004). "Crossover interference." Current Biology **14**(24): R1036-R1037.

Hilliker, A. J., S. H. Clark, et al. (1991). "The effect of DNA sequence polymorphisms on intragenic recombination in the rosy locus of Drosophila melanogaster." Genetics **129**(3): 779-781.

Hinch, A. G., A. Tandon, et al. (2011). "The landscape of recombination in African Americans." Nature **476**(7359): 170-U167.

Hochwagen, A. and G. A. Marais (2010). "Meiosis: a PRDM9 guide to the hotspots of recombination." Curr Biol **20**(6): R271-274.

Hunt, B. G., K. M. Glastad, et al. (2013). "Patterning and regulatory associations of DNA methylation are mirrored by histone modifications in insects." Genome Biology and Evolution **5**(3): 591-598.

Hunt, G. J. and R. E. Page (1995). "Linkage Map of the Honey-Bee, Apis-Mellifera, Based on Rapd Markers." Genetics **139**(3): 1371-1382.

Inoue, K. and J. R. Lupski (2002). "Molecular mechanisms for genomic disorders." Annu Rev Genomics Hum Genet **3**: 199-242.

Jeffreys, A. J., L. Kauppi, et al. (2001). "Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex." Nature Genetics **29**(2): 217-222.

Jeffreys, A. J., J. Murray, et al. (1998). "High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot." Molecular Cell **2**(2): 267-273.

Jeffreys, A. J. and R. Neumann (2002). "Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot." Nature Genetics **31**(3): 267-271.

Jeffreys, A. J. and R. Neumann (2005). "Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot." Human Molecular Genetics **14**(15): 2277-2287.

Jeffreys, A. J. and R. Neumann (2009). "The rise and fall of a human recombination hot spot." Nature Genetics **41**(5): 625-629.

Jeffreys, A. J., R. Neumann, et al. (2005). "Human recombination hot spots hidden in regions of strong marker association." Nature Genetics **37**(6): 601-606.

Jeffreys, A. J., A. Ritchie, et al. (2000). "High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot." Human Molecular Genetics **9**(5): 725-733.

Jeltsch, A., W. Nellen, et al. (2006). "Two substrates are better than one: dual specificities for Dnmt2 methyltransferases." Trends Biochem Sci **31**(6): 306-308.

Jensen-Seaman, M. I., T. S. Furey, et al. (2004). "Comparative recombination rates in the rat, mouse, and human genomes." Genome Research **14**(4): 528-538.

Jensen, J. D. and D. Bachtrog (2011). "Characterizing the influence of effective population size on the rate of adaptation: Gillespie's Darwin domain." Genome Biology and Evolution **3**: 687-701.

JH, G. (2004). Population genetics: a concise guide. Baltimore, The Johns Hopkins University Press.

Jordan, R. A. and R. W. Brosemer (1974). "Characterization of DNA from 3 Bee Species." Journal of Insect Physiology **20**(12): 2513-2520.

Jurkowski, T. P. and A. Jeltsch (2011). "On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2." Plos One **6**(11): e28104.

Jurkowski, T. P., M. Meusburger, et al. (2008). "Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism." RNA **14**(8): 1663-1670.

Kaneda, M., T. Sado, et al. (2004). "Role of de novo DNA methyltransferases in initiation of genomic imprinting and X-chromosome inactivation." Cold Spring Harb Symp Quant Biol **69**: 125-129.

Kaplan, N. L., R. R. Hudson, et al. (1989). "The "hitchhiking effect" revisited." Genetics **123**(4): 887-899.

Kaplan, T., N. Friedman, et al. (2005). "Ab initio prediction of transcription factor targets using structural knowledge." PLoS Comput Biol **1**(1): e1.

Kauppi, L., A. J. Jeffreys, et al. (2004). "Where the crossovers are: Recombination distributions in mammals." Nature Reviews Genetics **5**(6): 413-424.

Kauppi, L., M. P. H. Stumpf, et al. (2005). "Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human MHC class II region." Genomics **86**(1): 13-24.

Kaur, T. and M. V. Rockman (2014). "Crossover heterogeneity in the absence of hotspots in Caenorhabditis elegans." Genetics **196**(1): 137-148.

Keeney, S. (2001). "Mechanism and control of meiotic recombination initiation." Current Topics in Developmental Biology, Vol 52 **52**: 1-53.

Kelmenson, P. M., P. Petkov, et al. (2005). "A torrid zone on mouse chromosome 1 containing a cluster of recombinational hotspots." Genetics **169**(2): 833-841.

Kern, A. D., C. D. Jones, et al. (2002). "Genomic effects of nucleotide substitutions in Drosophila simulans." Genetics **162**(4): 1753-1761.

Khil, P. P. and R. D. Camerini-Otero (2010). "Genetic crossovers are predicted accurately by the computed human recombination map." PLoS Genet **6**(1): e1000831.

Kiani, J., V. Grandjean, et al. (2013). "RNA-mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2." PLoS Genet **9**(5): e1003498.

Kim, S., V. Plagnol, et al. (2007). "Recombination and linkage disequilibrium in Arabidopsis thaliana." Nature Genetics **39**(9): 1151-1155.

Kim, S., V. Plagnol, et al. (2007). "Recombination and linkage disequilibrium in Arabidopsis thaliana." Nat Genet **39**(9): 1151-1155.

Kim, Y. (2004). "Effect of strong directional selection on weakly selected mutations at linked sites: implication for synonymous codon usage." Molecular Biology and Evolution **21**(2): 286-294.

Kim, Y. and W. Stephan (2002). "Detecting a local signature of genetic hitchhiking along a recombining chromosome." Genetics **160**(2): 765-777.

Kimura, M. (1983). The neutral theory of molecular evolution. New York, Cambridge University Press.

Kondrashov, A. S. (1988). "Deleterious mutations and the evolution of sexual reproduction." Nature **336**(6198): 435-440.

Kong, A., D. F. Gudbjartsson, et al. (2002). "A high-resolution recombination map of the human genome." Nature Genetics **31**(3): 241-247.

Kong, A., G. Thorleifsson, et al. (2010). "Fine-scale recombination rate differences between sexes, populations and individuals." Nature **467**(7319): 1099-1103.

Korol, A. B. and K. G. Iliadi (1994). "Increased recombination frequencies resulting from directional selection for geotaxis in Drosophila." Heredity (Edinb) **72 (Pt 1)**: 64-68.

Kosambi, D. (1943). "The estimation of map distances from recombination values." Annals of Human Genetics **12**: 172–175.

Kraft, T., T. Sall, et al. (1998). "Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (Beta vulgaris subsp. maritima)." Genetics **150**(3): 1239-1244.

Krauss, V. and G. Reuter (2011). "DNA Methylation in Drosophila-A Critical Evaluation." Modifications of Nuclear DNA and Its Regulatory Proteins **101**: 177-191.

Kulathinal, R. J., S. M. Bennett, et al. (2008). "Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence." Proc Natl Acad Sci U S A **105**(29): 10051-10056.

Kulathinal, R. J., L. S. Stevison, et al. (2009). "The genomics of speciation in Drosophila: diversity, divergence, and introgression estimated using low-coverage genome sequencing." PLoS Genet **5**(7): e1000550.

Lake, C. M., R. J. Nielsen, et al. (2011). "The Drosophila Zinc Finger Protein Trade Embargo Is Required for Double Strand Break Formation in Meiosis." Plos Genetics **7**(2).

Langley, C. H., B. P. Lazzaro, et al. (2000). "Linkage disequilibria and the site frequency spectra in the su(s) and su(w(a)) regions of the Drosophila melanogaster X chromosome." Genetics **156**(4): 1837-1852.

Larracuente, A. M., T. B. Sackton, et al. (2008). "Evolution of protein-coding genes in Drosophila." Trends Genet **24**(3): 114-123.

Lenormand, T. and S. P. Otto (2000). "The evolution of recombination in a heterogeneous environment." Genetics **156**(1): 423-438.

Lercher, M. J. and L. D. Hurst (2002). "Human SNP variability and mutation rate are higher in regions of high recombination." Trends in Genetics **18**(7): 337-340.

Li, H., B. Handsaker, et al. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, J., R. A. Harris, et al. (2012). "Genomic hypomethylation in the human germline associates with selective structural mutability in the human genome." PLoS Genet **8**(5): e1002692.

Lichten, M. (2008). Meiotic chromatin: the substrate for recombination initiation. Recombination and Meiosis: Models, Means, and Evolution. D. H. L. R. Egel. Berlin, Springer-Verlag**:** 165-193.

Lichten, M., R. H. Borts, et al. (1987). "Meiotic Gene Conversion and Crossing over between Dispersed Homologous Sequences Occurs Frequently in Saccharomyces-Cerevisiae." Genetics **115**(2): 233-246.

Lichten, M. and A. S. H. Goldman (1995). "Meiotic recombination hotspots." Annual Review of Genetics **29**: 423-444.

Lin, M. J., L. Y. Tang, et al. (2005). "DNA methyltransferase gene dDnmt2 and longevity of Drosophila." J Biol Chem **280**(2): 861-864.

Loewe, L. and B. Charlesworth (2007). "Background selection in single genes may explain patterns of codon bias." Genetics **175**(3): 1381-1393.

Loewe, L., B. Charlesworth, et al. (2006). "Estimating selection on nonsynonymous mutations." Genetics **172**(2): 1079-1092.

Loewe, L. and W. G. Hill (2010). "The population genetics of mutations: good, bad and indifferent." Philos Trans R Soc Lond B Biol Sci **365**(1544): 1153-1167.

Lohmueller, K. E., A. Albrechtsen, et al. (2011). "Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome." PLoS Genet **7**(10): e1002326.

Lyko, F. (2001). "DNA methylation learns to fly." Trends Genet **17**(4): 169-172.

Lyko, F., C. Beisel, et al. (2006). "Epigenetic regulation in Drosophila." DNA Methylation: Development, Genetic Disease and Cancer **310**: 23-44.

Lyko, F., N. Kunert, et al. (2003). "A Dnmt2-1ike protein mediates DNA methylation in Drosophila." Development **130**(21): 5083-5090.

Lyko, F., F. Weissmann, et al. (2003). "DNA hypermethylation in Drosophila melanogaster causes irregular chromosome condensation and dysregulation of epigenetic histone modifications." Molecular and Cellular Biology **23**(7): 2577-2586.

Lyko, F., A. J. Whittaker, et al. (2000). "The putative Drosophila methyltransferase gene dDnmt2 is contained in a transposon-like element and is expressed specifically in ovaries." Mech Dev **95**(1-2): 215-217.

Lyne, R., R. Smith, et al. (2007). "FlyMine: an integrated database for Drosophila and Anopheles genomics." Genome Biology **8**(7).

Machado, C. A., T. S. Haselkorn, et al. (2007). "Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in Drosophila pseudoobscura and D. persimilis." Genetics **175**(3): 1289-1306.

Machado, C. A., R. M. Kliman, et al. (2002). "Inferring the history of speciation from multilocus DNA sequence data: the case of Drosophila pseudoobscura and close relatives." Molecular Biology and Evolution **19**(4): 472-488.

Mackay, T. F., S. Richards, et al. (2012). "The Drosophila melanogaster Genetic Reference Panel." Nature **482**(7384): 173-178.

Macpherson, J. M., G. Sella, et al. (2007). "Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in Drosophila." Genetics **177**(4): 2083-2099.

Main, B. J., A. D. Smith, et al. (2013). "Transcription start site evolution in Drosophila." Molecular Biology and Evolution **30**(8): 1966-1974.

Maloisel, L. and J. L. Rossignol (1998). "Suppression of crossing-over by DNA methylation in Ascobolus." Genes Dev **12**(9): 1381-1389.

Mancera, E., R. Bourgon, et al. (2008). "High-resolution mapping of meiotic crossovers and non-crossovers in yeast." Nature **454**(7203): 479-U471.

Manzano-Winkler, B., S. E. McGaugh, et al. (2013). "How hot are drosophila hotspots? examining recombination rate variation and associations with nucleotide diversity, divergence, and maternal age in Drosophila pseudoobscura." Plos One **8**(8): e71582.

Marais, G. (2003). "Biased gene conversion: implications for genome and sex evolution." Trends in Genetics **19**(6): 330-338.

Marais, G., D. Mouchiroud, et al. (2001). "Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes." Proceedings of the National Academy of Sciences of the United States of America **98**(10): 5688-5692.

Margarido, G. R., A. P. Souza, et al. (2007). "OneMap: software for genetic mapping in outcrossing species." Hereditas **144**(3): 78-79.

Marhold, J., A. Brehm, et al. (2004). "The Drosophila methyl-DNA binding protein MBD2/3 interacts with the NuRD complex via p55 and MI-2." Bmc Molecular Biology **5**.

Marhold, J., N. Rothe, et al. (2004). "Conservation of DNA methylation in dipteran insects." Insect Mol Biol **13**(2): 117-123.

Martin, G., S. P. Otto, et al. (2006). "Selection for recombination in structured populations." Genetics **172**(1): 593-609.

McGaugh, S. E., C. S. Heil, et al. (2012). "Recombination modulates how selection affects linked sites in Drosophila." PLoS Biol **10**(11): e1001422.

McGaugh, S. E. and M. A. Noor (2012). "Genomic impacts of chromosomal inversions in parapatric Drosophila species." Philos Trans R Soc Lond B Biol Sci **367**(1587): 422-429.

McGaugh, S. E., C. S. Smukowski, et al. (2012). "Recombination modulates how selection affects linked sites in Drosophila." Nature Precedings

McKenna, A., M. Hanna, et al. (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

McKim, K. S., J. K. Jang, et al. (2002). "Meiotic recombination and chromosome segregation in Drosophila females." Annu Rev Genet **36**: 205-232.

McQuilton, P., S. E. St Pierre, et al. (2012). "FlyBase 101-the basics of navigating FlyBase." Nucleic Acids Research **40**(D1): D706-D714.

McVean, G., P. Awadalla, et al. (2002). "A coalescent-based method for detecting and estimating recombination from gene sequences." Genetics **160**(3): 1231-1241.

McVean, G. A. and B. Charlesworth (2000). "The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation." Genetics **155**(2): 929-944.

McVean, G. A., S. R. Myers, et al. (2004). "The fine-scale structure of recombination rate variation in the human genome." Science **304**(5670): 581-584.

Melamed-Bessudo, C. and A. A. Levy (2012). "Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis." Proc Natl Acad Sci U S A **109**(16): E981-988.

Meunier, J. and L. Duret (2004). "Recombination drives the evolution of GC-content in the human genome." Molecular Biology and Evolution **21**(6): 984-990.

Meznar, E. R., J. Gadau, et al. (2010). "Comparative Linkage Mapping Suggests a High Recombination Rate in All Honeybees." Journal of Heredity **101**: S118-S126.

Miller, D. E., S. Takeo, et al. (2012). "A whole-chromosome analysis of meiotic recombination in Drosophila melanogaster." G3: Genes, Genomes, Genetics **2**(2): 249-260.

Mirouze, M., M. Lieberman-Lazarovich, et al. (2012). "Loss of DNA methylation affects the recombination landscape in Arabidopsis." Proc Natl Acad Sci U S A **109**(15): 5880-5885.

Montgomery, E. A., S. M. Huang, et al. (1991). "Chromosome Rearrangement by Ectopic Recombination in Drosophila-Melanogaster - Genome Structure and Evolution." Genetics **129**(4): 1085-1098.

Morgan, T. H. (1910). "Sex limited inheritance in Drosophila." Science **32**(1): 120-122.

Morran, L. T., M. D. Parmenter, et al. (2009). "Mutation load and rapid adaptation favour outcrossing over self-fertilization." Nature **462**(7271): 350-352.

Munoz-Fuentes, V., A. Di Rienzo, et al. (2011). "Prdm9, a Major Determinant of Meiotic Recombination Hotspots, Is Not Functional in Dogs and Their Wild Relatives, Wolves and Coyotes." Plos One **6**(11).

Myers, S., L. Bottolo, et al. (2005). "A fine-scale map of recombination rates and hotspots across the human genome." Science **310**(5746): 321-324.

Myers, S., R. Bowden, et al. (2010). "Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination." Science **327**(5967): 876-879.

Myers, S., R. Bowden, et al. (2010). "Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination." Science **327**(5967): 876-879.

Myers, S., C. Freeman, et al. (2008). "A common sequence motif associated with recombination hot spots and genome instability in humans." Nat Genet **40**(9): 1124-1129.

Myers, S., C. C. A. Spencer, et al. (2006). "The distribution and causes of meiotic recombination in the human genome." Biochemical Society Transactions **34**: 526-530.

Nachman, M. W. (1997). "Patterns of DNA variability at X-linked loci in Mus domesticus." Genetics **147**(3): 1303-1316.

Nachman, M. W. (2002). "Variation in recombination rate across the genome: evidence and implications." Current Opinion in Genetics & Development **12**(6): 657-663.

Nachman, M. W., V. L. Bauer, et al. (1998). "DNA variability and recombination rates at X-linked loci in humans." Genetics **150**(3): 1133-1141.

Nanty, L., G. Carbajosa, et al. (2011). "Comparative methylomics reveals gene-body H3K36me3 in Drosophila predicts DNA methylation and CpG landscapes in other invertebrates." Genome Res **21**(11): 1841-1850.

Neumann, R. and A. J. Jeffreys (2006). "Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation." Human Molecular Genetics **15**(9): 1401-1411.

Neumann, R. and A. J. Jeffreys (2006). "Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation." Hum Mol Genet **15**(9): 1401-1411.

Noor, M. A., D. A. Garfield, et al. (2007). "Divergence between the Drosophila pseudoobscura and D. persimilis genome sequences in relation to chromosomal inversions." Genetics **177**(3): 1417-1428.

Noor, M. A., M. D. Schug, et al. (2000). "Microsatellite variation in populations of Drosophila pseudoobscura and Drosophila persimilis." Genetical Research **75**(1): 25-35.

Noor, M. A. F. (2008). "Connecting recombination, nucleotide diversity and species divergence in Drosophila." Fly **2**(5): -.

Noor, M. A. F. (2008). "Mutagenesis from Meiotic Recombination Is Not a Primary Driver of Sequence Divergence between Saccharomyces Species." Molecular Biology and Evolution **25**(11): 2439-2444.

Nordborg, M., B. Charlesworth, et al. (1996). "Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species." Proceedings of the Royal Society of London Series B-Biological Sciences **263**(1373): 1033-1039.

Nordborg, M., T. T. Hu, et al. (2005). "The pattern of polymorphism in Arabidopsis thaliana." PLoS Biol **3**(7): e196.

Nowell, R. W., B. Charlesworth, et al. (2011). "Ancestral polymorphisms in Drosophila pseudoobscura and Drosophila miranda." Genet Res (Camb) **93**(4): 255-263.

Okano, M., S. Xie, et al. (1998). "Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells." Nucleic Acids Research **26**(11): 2536-2540.

Oliver, P. L., L. Goodstadt, et al. (2009). "Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa." Plos Genetics **5**(12): -.

Ortiz-Barrientos, D., S. C. Audrey, et al. (2006). "A recombinational portrait of the Drosophila pseudoobscura genome." Genetical Research **87**(1): 23-31.

Otto, S. P. and N. H. Barton (2001). "Selection for recombination in small populations." Evolution **55**(10): 1921-1931.

Otto, S. P. and T. Lenormand (2002). "Resolving the paradox of sex and recombination." Nat Rev Genet **3**(4): 252-261.

Otto, S. P. and Y. Michalakis (1998). "The evolution of recombination in changing environments." Trends in Ecology & Evolution **13**(4): 145-151.

Page, S. L. and R. S. Hawley (2003). "Chromosome choreography: The meiotic ballet." Science **301**(5634): 785-789.

Paigen, K., J. P. Szatkiewicz, et al. (2008). "The Recombinational Anatomy of a Mouse Chromosome." Plos Genetics **4**(7): -.

Pal, C., B. Papp, et al. (2001). "Does the recombination rate affect the efficiency of purifying selection? The yeast genome provides a partial answer." Molecular Biology and Evolution **18**(12): 2323-2326.

Palsson, A., A. Rouse, et al. (2004). "Nucleotide variation in the Egfr locus of Drosophila melanogaster." Genetics **167**(3): 1199-1212.

Pan, J., M. Sasaki, et al. (2011). "A Hierarchical Combination of Factors Shapes the Genome-wide Topography of Yeast Meiotic Recombination Initiation." Cell **144**(5): 719-731.

Papavasiliou, F. N. and D. G. Schatz (2000). "Cell-cycle-regulated DNA double-strand breaks in somatic hypermutation of immunoglobulin genes." Nature **408**(6809): 216-221.

Parvanov, E. D., P. M. Petkov, et al. (2010). "Prdm9 controls activation of mammalian recombination hotspots." Science **327**(5967): 835.

Parvanov, E. D., P. M. Petkov, et al. (2010). "Prdm9 Controls Activation of Mammalian Recombination Hotspots." Science **327**(5967): 835-835.

Payseur, B. A. and M. W. Nachman (2000). "Microsatellite variation and recombination rate in the human genome." Genetics **156**(3): 1285-1298.

Petes, T. D. (2001). "Meiotic recombination hot spots and cold spots." Nature Reviews Genetics **2**(5): 360-369.

Phalke, S., O. Nickel, et al. (2009). "Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2." Nat Genet **41**(6): 696-702.

Ponting, C. P. (2011). "What are the genomic drivers of the rapid evolution of PRDM9?" Trends in Genetics **27**(5): 165-171.

Powell, J. R. (1983). "Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from Drosophila." Proc Natl Acad Sci U S A **80**(2): 492-495.

Presgraves, D. C. (2005). "Recombination enhances protein adaptation in Drosophila melanogaster." Curr Biol **15**(18): 1651-1656.

Ptak, S. E., D. A. Hinds, et al. (2005). "Fine-scale recombination patterns differ between chimpanzees and humans." Nature Genetics **37**(4): 429-434.

Ptak, S. E., A. D. Roeder, et al. (2004). "Absence of the TAP2 human recombination hotspot in chimpanzees." Plos Biology **2**(6): 849-855.

Raddatz, G., P. M. Guzzardo, et al. (2013). "Dnmt2-dependent methylomes lack defined DNA methylation patterns." Proc Natl Acad Sci U S A **110**(21): 8627-8631.

Radford, S. J., M. M. Sabourin, et al. (2007). "Meiotic recombination in Drosophila msh6 mutants yields discontinuous gene conversion tracts." Genetics **176**(1): 53-62.

Rasmussen, S. W. (1977). "Transformation of Synaptonemal Complex into Elimination Chromatin in Bombyx-Mori Oocytes." Chromosoma **60**(3): 205-221.

Rattray, A. J., C. B. McGill, et al. (2001). "Fidelity of mitotic double-strand-break repair in Saccharomyces cerevisiae: A role for SAE2/COM1." Genetics **158**(1): 109-122.

Reed, F. A. and S. A. Tishkoff (2006). "Positive selection can create false hotspots of recombination." Genetics **172**(3): 2011-2014.

Rees, H. and P. J. Dale (1974). "Chiasmata and Variability in Lolium and Festuca Populations." Chromosoma **47**(3): 335-351.

Rice, P., I. Longden, et al. (2000). "EMBOSS: The European molecular biology open software suite." Trends in Genetics **16**(6): 276-277.

Rice, W. R. (1989). "Analyzing Tables of Statistical Tests." Evolution **43**(1): 223-225.

Richards, S., Y. Liu, et al. (2005). "Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution." Genome Res **15**(1): 1-18.

Robertson, K. D. and A. P. Wolffe (2000). "DNA methylation in health and disease." Nat Rev Genet **1**(1): 11-19.

Rockman, M. V., S. S. Skrovanek, et al. (2010). "Selection at linked sites shapes heritable phenotypic variation in C. elegans." Science **330**(6002): 372-376.

Roder, K., M. S. Hung, et al. (2000). "Transcriptional repression by Drosophila methyl-CpG-binding proteins." Mol Cell Biol **20**(19): 7401-7409.

Roeder, G. S. (1997). "Meiotic chromosomes: it takes two to tango." Genes & Development **11**(20): 2600-2621.

Roselius, K., W. Stephan, et al. (2005). "The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species." Genetics **171**(2): 753-763.

Ross-Ibarra, J. (2004). "The evolution of recombination under domestication: A test of two hypotheses." American Naturalist **163**(1): 105-112.

Rossignol, J. L. and G. Faugeron (1994). "Gene inactivation triggered by recognition between DNA repeats." Experientia **50**(3): 307-317.

Rouyer, F., M. C. Simmler, et al. (1987). "A Sex-Chromosome Rearrangement in a Human Xx Male Caused by Alu-Alu Recombination." Cell **51**(3): 417-425.

Salzberg, A., O. Fisher, et al. (2004). "Identification of methylated sequences in genomic DNA of adult Drosophila melanogaster." Biochem Biophys Res Commun **322**(2): 465-469.

Sapienza, P.-M. d. V. F. a. C. (2001). "Recombination is proportional to the number of chromosome arms in mammals." Mammalian Genome **12**: 318-322.

Sattath, S., E. Elyashiv, et al. (2011). "Pervasive Adaptive Protein Evolution Apparent in Diversity Patterns around Amino Acid Substitutions in Drosophila simulans." Plos Genetics **7**(2).

Sattath, S., E. Elyashiv, et al. (2011). "Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in Drosophila simulans." PLoS Genet **7**(2): e1001302.

Schaefer, M. and F. Lyko (2010). "Solving the Dnmt2 enigma." Chromosoma **119**(1): 35-40.

Schaefer, M., T. Pollex, et al. (2010). "RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage." Genes Dev **24**(15): 1590-1595.

Schaeffer, S. W. and E. L. Miller (1992). "Molecular population genetics of an electrophoretically monomorphic protein in the alcohol dehydrogenase region of Drosophila pseudoobscura." Genetics **132**(1): 163-178.

Schaeffer, S. W., C. S. Walthour, et al. (2001). "Protein variation in Adh and Adh-related in Drosophila pseudoobscura. Linkage disequilibrium between single nucleotide polymorphisms and protein alleles." Genetics **159**(2): 673-687.

Schmid, M., I. Nanda, et al. (2005). "Second report on chicken genes and chromosomes 2005." Cytogenetic and Genome Research **109**(4): 415-479.

Schultz, J. and H. Redfield (1951). "Interchromosomal Effects on Crossing over in Drosophila." Cold Spring Harbor Symposia on Quantitative Biology **16**: 175-197.

Schurko, A. M. and J. M. Logsdon (2008). "Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex." Bioessays **30**(6): 579-589.

Schweitzer, M. D. (1935). "An Analytical Study of Crossing over in Drosophila Melanogaster." Genetics **20**(5): 497-527.

Segurel, L., E. M. Leffler, et al. (2011). "The Case of the Fickle Fingers: How the PRDM9 Zinc Finger Protein Specifies Meiotic Recombination Hotspots in Humans." Plos Biology **9**(12).

Sella, G., D. A. Petrov, et al. (2009). "Pervasive Natural Selection in the Drosophila Genome?" Plos Genetics **5**(6): -.

Shifman, S., J. T. Bell, et al. (2006). "A high-resolution single nucleotide polymorphism genetic map of the mouse genome." Plos Biology **4**(12): 2227-2237.

Sigurdsson, M. I., A. V. Smith, et al. (2009). "HapMap methylation-associated SNPs, markers of germline DNA methylation, positively correlate with regional levels of human meiotic recombination." Genome Res **19**(4): 581-589.

Singh, N. D., C. F. Aquadro, et al. (2009). "Estimation of Fine-Scale Recombination Intensity Variation in the white-echinus Interval of D-melanogaster." Journal of Molecular Evolution **69**(1): 42-53.

Singh, N. D., E. A. Stone, et al. (2013). "Fine-scale heterogeneity in crossover rate in the garnet-scalloped region of the Drosophila melanogaster X chromosome." Genetics **194**(2): 375-387.

Slatkin, M. (2008). "Linkage disequilibrium - understanding the evolutionary past and mapping the medical future." Nature Reviews Genetics **9**(6): 477-485.

Slotte, T., T. Bataillon, et al. (2011). "Genomic determinants of protein evolution and polymorphism in Arabidopsis." Genome Biology and Evolution **3**: 1210-1219.

Smagulova, F., I. V. Gregoretti, et al. (2011). "Genome-wide analysis reveals novel molecular features of mouse recombination hotspots." Nature **472**(7343): 375-378.

Small, K., J. Iber, et al. (1997). "Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats." Nat Genet **16**(1): 96-99.

Smith, J. M. and J. Haigh (1974). "Hitch-Hiking Effect of a Favorable Gene." Genetical Research **23**(1): 23-35.

Smukowski, C. S. and M. A. F. Noor (2011). "Recombination rate variation in closely related species." Heredity **107**(6): 496-508.

Solignac, M., D. Vautrin, et al. (2004). "A microsatellite-based linkage map of the Honeybee, Apis mellifera L." Genetics **167**(1): 253-262.

Sommermeyer, V., C. Beneut, et al. (2013). "Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes." Mol Cell **49**(1): 43-54.

Spencer, C. C. A., P. Deloukas, et al. (2006). "The influence of recombination on human genetic diversity." Plos Genetics **2**(9): 1375-1385.

St Pierre, S. E., L. Ponting, et al. (2014). "FlyBase 102--advanced approaches to interrogating FlyBase." Nucleic Acids Research **42**(1): D780-788.

Stahl, F. W., H. M. Foss, et al. (2004). "Does crossover interference count in Saccharomyces cerevisiae?" Genetics **168**(1): 35-48.

Steiner, W. W., P. A. Davidow, et al. (2011). "Important Characteristics of Sequence-Specific Recombination Hotspots in Schizosaccharomyces pombe." Genetics **187**(2): 385-396.

Steiner, W. W., E. M. Steiner, et al. (2009). "Novel Nucleotide Sequence Motifs That Produce Hotspots of Meiotic Recombination in Schizosaccharomyces pombe." Genetics **182**(2): 459-469.

Stephan, W. (2010). "Genetic hitchhiking versus background selection: the controversy and its implications." Philos Trans R Soc Lond B Biol Sci **365**(1544): 1245-1253.

Stephan, W. and C. H. Langley (1989). "Molecular genetic variation in the centromeric region of the X chromosome in three Drosophila ananassae populations. I. Contrasts between the vermilion and forked loci." Genetics **121**(1): 89-99.

Stephan, W. and C. H. Langley (1998). "DNA polymorphism in Lycopersicon and crossing-over per physical length." Genetics **150**(4): 1585-1593.

Stephan, W., Wiehe THE, Lenz MW (1992). "The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory." Theoretical Population Biology **41**: 237-254.

Stevison, L. S., and M.A.F. Noor (2009). Recombination rates in Drosophila. Encyclopedia of Life Sciences. Chichester, John Wiley & Sons, Ltd.

Stevison, L. S. and M. A. F. Noor (2010). "Genetic and Evolutionary Correlates of Fine-Scale Recombination Rate Variation in Drosophila persimilis." Journal of Molecular Evolution **71**(5-6): 332-345.

Strathern, J. N., B. K. Shafer, et al. (1995). "DNA-Synthesis Errors Associated with Double-Strand-Break Repair." Genetics **140**(3): 965-972.

Tajima, F. (1990). "Relationship between DNA polymorphism and fixation time." Genetics **125**(2): 447-454.

Takahashi, A., Y. H. Liu, et al. (2004). "Genetic variation versus recombination rate in a structured population of mice." Molecular Biology and Evolution **21**(2): 404-409.

Tamura, K., S. Subramanian, et al. (2004). "Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks." Molecular Biology and Evolution **21**(1): 36-44.

Tan, Y. D. and Y. X. Fu (2006). "A novel method for estimating linkage maps." Genetics **173**(4): 2383-2390.

Tenaillon, M. I., M. C. Sawkins, et al. (2002). "Patterns of diversity and recombination along chromosome 1 of maize (Zea mays ssp mays L.)." Genetics **162**(3): 1401-1413.

Tenaillon, M. I., M. C. Sawkins, et al. (2001). "Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp mays L.)." Proceedings of the National Academy of Sciences of the United States of America **98**(16): 9161-9166.

Tenaillon, M. I., J. U'Ren, et al. (2004). "Selection versus demography: A multilocus investigation of the domestication process in maize." Molecular Biology and Evolution **21**(7): 1214-1225.

Thuriaux, P. (1977). "Is Recombination Confined to Structural Genes on Eukaryotic Genome." Nature **268**(5619): 460-462.

True, J. R., J. M. Mercer, et al. (1996). "Differences in crossover frequency and distribution among three sibling species of Drosophila." Genetics **142**(2): 507-523.

Tsai, I. J., A. Burt, et al. (2010). "Conservation of recombination hotspots in yeast." Proceedings of the National Academy of Sciences of the United States of America **107**(17): 7847-7852.

Tuorto, F., R. Liebers, et al. (2012). "RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis." Nat Struct Mol Biol **19**(9): 900-905.

Tweedie, S., H. H. Ng, et al. (1999). "Vestiges of a DNA methylation system in Drosophila melanogaster?" Nat Genet **23**(4): 389-390.

Van Os, H., P. Stam, et al. (2005). "RECORD: a novel method for ordering loci on a genetic linkage map." Theoretical and Applied Genetics **112**(1): 30-40.

Vicario, S., E. N. Moriyama, et al. (2007). "Codon usage in twelve species of Drosophila." BMC Evol Biol **7**: 226.

Vicoso, B. and B. Charlesworth (2006). "Evolution on the X chromosome: unusual patterns and processes." Nat Rev Genet **7**(8): 645-653.

Warton, D. I. and F. K. Hui (2011). "The arcsine is asinine: the analysis of proportions in ecology." Ecology **92**(1): 3-10.

Weber, C. C. and L. D. Hurst (2009). "Protein rates of evolution are predicted by double-strand break events, independent of crossing-over rates." Genome Biology and Evolution **1**: 340-349.

Webster, M. T. and L. D. Hurst (2012). "Direct and indirect consequences of meiotic recombination: implications for genome evolution." Trends Genet **28**(3): 101-109.

Whitkus, R., J. Doebley, et al. (1992). "Comparative genome mapping of Sorghum and maize." Genetics **132**(4): 1119-1130.

Wilfert, L., J. Gadau, et al. (2007). "Variation in genomic recombination rates among animal taxa and the case of social insects." Heredity **98**(4): 189-197.

Wilson, K., and ICW Hardy (2002). Statistical analysis of sex ratios: an introduction. Sex ratios: concepts and research methods. I. Hardy. New York, Cambridge University Press**:** 48-92.

Winckler, W., S. R. Myers, et al. (2005). "Comparison of fine-scale recombination rates in humans and chimpanzees." Science **308**(5718): 107-111.

Wolfe, S. A., L. Nekludova, et al. (2000). "DNA recognition by Cys(2)His(2) zinc finger proteins." Annual Review of Biophysics and Biomolecular Structure **29**: 183-212.

Wong, A. K., A. L. Ruhe, et al. (2010). "A Comprehensive Linkage Map of the Dog Genome." Genetics **184**(2): 595-U436.

Workman, C. T., Y. T. Yin, et al. (2005). "enoLOGOS: a versatile web tool for energy normalized sequence logos." Nucleic Acids Research **33**: W389-W392.

Yauk, C. L., P. R. J. Bois, et al. (2003). "High-resolution sperm typing of meiotic recombination in the mouse MHC E-beta gene." Embo Journal **22**(6): 1389-1397.

Yelina, N. E., K. Choi, et al. (2012). "Epigenetic remodeling of meiotic crossover frequency in Arabidopsis thaliana DNA methyltransferase mutants." PLoS Genet **8**(8): e1002844.

Yi, S., T. J. Summers, et al. (2004). "Recombination has little effect on the rate of sequence divergence in pseudoautosomal boundary 1 among humans and great apes." Genome Research **14**(1): 37-43.

Yi, S. J. and W. H. Li (2005). "Molecular evolution of recombination hotspots and highly recombining pseudoautosomal regions in hominoids." Molecular Biology and Evolution **22**(5): 1223-1230.

Yoder, J. A., C. P. Walsh, et al. (1997). "Cytosine methylation and the ecology of intragenomic parasites." Trends in Genetics **13**(8): 335-340.

Zemach, A., I. E. McDaniel, et al. (2010). "Genome-wide evolutionary analysis of eukaryotic DNA methylation." Science **328**(5980): 916-919.

Zemach, A. and D. Zilberman (2010). "Evolution of eukaryotic DNA methylation and the pursuit of safer sex." Curr Biol **20**(17): R780-785.

Biography

Caitlin Smukowski Heil was born in Bellevue, Washington on October 24, 1986 to David and Patty Smukowski. After attending Seattle Preparatory High School, she graduated *cum laude* with a Bachelor of Arts in Biology from the University of Pennsylvania in 2009. Caiti became interested in evolutionary genetics in her senior year, completing an honors thesis entitled "An analysis of candidate interaction genes with couch potato, a gene for reproductive diapauses in *Drosophila melanogaster*" under the guidance of Paul Schmidt. In the fall of 2009, Caiti joined Mohamed Noor's lab at Duke University, where she has published five peer-reviewed articles ("Zinc finger binding motifs do not explain recombination variation within or between species of Drosophila"; "Recombination modulates how selection affects linked sites in Drosophila"; "Recombination rate variation in closely related species"; "Witnessing Evolution First-hand: A K-12 Laboratory Exercise in Genetics and Evolution Using Drosophila"; and "Witnessing phenotypic and molecular evolution in the fruit fly"), and two non-peer reviewed papers ("Studying recombination with high-throughput sequencing: An educational primer for use with 'Fine-scale heterogeneity in crossover rate in the garnet-scalloped region of the Drosophila melanogaster X chromosome'"; and "Mentor vs. Monolith: Finding and being a good graduate advisor"). Caiti is a recipient of the James B. Duke fellowship and the National Science Foundation Graduate Fellowship, and has received grants from Sigma Xi, Society for the Study of Evolution, and the National Science

Foundation (DDIG). She is continuing on to a Postdoctoral position in Maitreya Dunham's lab at the University of Washington in 2014, where she will focus on genome evolution. She lives with her husband, Lou, and two cats, Monkey and Business, and enjoys reading, cooking, running, and hiking.